ANL-6820

ANL-6820

Argonne National Laboratory

ASYMPTOTIC SOLUTIONS TO COMPOUND DECISION PROBLEMS

by

John R. Van Ryzin

LEGAL NOTICE

Price **$2.25** . *Available from the Office of Technical Services, Department of Commerce, Washington 25, D.C.*

ANL-6820
Mathematics and Computers
(TID-4500, 27th Ed.)
AEC Research and
Development Report

ARGONNE NATIONAL LABORATORY
9700 South Cass Avenue
Argonne, Illinois 60440

ASYMPTOTIC SOLUTIONS TO COMPOUND DECISION PROBLEMS

by

John R. Van Ryzin
Applied Mathematics Division*

February 1964

*A major part of this work was performed while the author was at Michigan State University.

Reproduced from a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Michigan State University.

Operated by The University of Chicago
under
Contract W-31-109-eng-38
with the
U. S. Atomic Energy Commission

TABLE OF CONTENTS

		Page
LIST OF APPENDICES		iii
ABSTRACT		v
INTRODUCTION		1
CHAPTER		
I.	THE FINITE COMPOUND DECISION PROBLEM	5
	1.1 Statement of the Problem	5
	1.2 Decision Procedures	6
	1.3 Estimation of Empirical Distributions on Ω	10
	1.4 Non-simple Decision Functions	14
II.	ASYMPTOTIC RESULTS FOR THE COMPOUND TESTING PROBLEM FOR TWO COMPLETELY SPECIFIED DISTRIBUTIONS	18
	2.1 Introduction and Notation	18
	2.2 An Inequality for the Regret Risk Function	21
	2.3 A Convergence Theorem of $0(N^{-1/2})$	23
	2.4 Convergence Theorems of Higher Order	27
	2.5 Examples Satisfying Theorem 3 or 4	32
III.	CONVERGENCE THEOREMS FOR THE GENERAL FINITE COMPOUND DECISION PROBLEM	38
	3.1 Introduction	38
	3.2 Uniform Convergence Theorem of $0(N^{-1/2})$	39
	3.3 Sufficient Conditions for a Theorem of Higher Order	41
	3.4 Uniform Convergence Theorem of $0(N^{-1})$	46
	3.5 Counter-example to Theorem 6 when (C) is Violated	50

TABLE OF CONTENTS

Page

IV. THE TWO-DECISION COMPOUND TESTING PROBLEM IN THE PRESENCE OF A NUISANCE PARAMETER 53

4.1 Introduction. 53

4.2 A Convergence Theorem in the Presence of a Nuisance Parameter 57

4.3 Examples for Theorem 7 64

4.4 Uniform Theorems in the Parameter τ. 69

4.5 Specific Results when s = 1 78

SUMMARY 83

APPENDICES 85

BIBLIOGRAPHY. 92

ACKNOWLEDGEMENTS 94

LIST OF APPENDICES

Page

APPENDIX

1. Proof that Condition (II") Implies Condition (II') when μ = P. 85

2. Truncation of $\overline{k}$ to a Convex Set of R^s 87

3. Extension of Results for a Randomized Procedure 89

ABSTRACT

ASYMPTOTIC SOLUTIONS TO
COMPOUND DECISION PROBLEMS

by John R. Van Ryzin

Simultaneous consideration of a large number of statistical decisions having identical generic structure constitutes a compound decision problem. In this thesis, decision procedures depending on data from all problems are shown to have certain optimal properties asymptotically as the number of problems increases.

More specifically, let X_α, $\alpha = 1,2,\ldots$ be a sequence of independent random variables with X_α having distribution P_{θ_α}, where θ_α takes a value in the finite parameter space $\Omega = \{0,\ldots,m-1\}$. Let the space of all sequences $\{\theta_\alpha, \alpha = 1,2,\ldots\}$ be denoted by Ω_∞. Fix N and consider the first N members of the sequence of X_α's. For each $\alpha = 1,\ldots,N$, it is required to make a decision d_α among n available decisions $\{0,\ldots,n-1\}$. Such an N-fold decision problem is called a finite compound decision problem.

Any N x n matrix of functions $T(x) = (t_{\alpha j}(x))$, where $t_{\alpha j} = \Pr\{d_\alpha = j | x\}$ with $x = (x_1,\ldots,x_N)$, $\alpha = 1,\ldots,N$, $j = 0,\ldots,n-1$, is a decision procedure for the N-fold compound problem. Define the risk of any such procedure, denoted by $R(\theta,T)$ for $\theta \in \Omega_\infty$, as the average of the risks for the N problems. With $p_i(\theta)$ as the relative frequency of problems in the first N problems having P_i as the governing distribution, $i = 0,\ldots,m-1$, we see that $p(\theta) = (p_0(\theta),\ldots,p_{m-1}(\theta))$ constitutes an empirical distribution on Ω. There exists a non-randomized procedure $t'_{p(\theta)}$ Bayes against $p(\theta)$ which has risk

$\phi(p(\theta)) = R(\theta, t'_{p(\theta)})$. The function $R(\theta,T) - \phi(p(\theta))$, called the regret risk function for the procedure T, is used as a measure of the optimality of the procedure T.

Existence of asymptotically good, unbiased estimates $\bar{h} = N^{-1} \sum_{\alpha=1}^{N} h(X_\alpha)$ of $p(\theta)$ is verified. To obtain procedures whose regret risk function converges to zero as $N \to \infty$, these estimates are substituted into the procedure $t'_{p(\theta)}$ to form the procedure $t'_{\bar{h}}$, which depends on data from all N problems. Under integrability assumptions on the kernel function h, convergence theorems for the regret risk function of $t'_{\bar{h}}$ are proved. These theorems are all uniform in $\theta \in \Omega_\infty$.

The main result is that if $|h|^3$ is integrable with respect to P_i, $i = 0,\ldots,m-1$, then the regret risk function of $t'_{\bar{h}}$ converges to zero at rate $O(N^{-1/2})$ uniformly in $\theta \in \Omega_\infty$. If $m = n = 2$, faster uniform convergence rates of $o(N^{-1/2})$ and $O(N^{-1})$ are attained under successively stronger continuity restrictions on P_0 and P_1 and integrability assumptions on h. A uniform theorem of $O(N^{-1})$ for the general m x n problem is also given under a strong continuity condition on the family $\{P_0,\ldots,P_{m-1}\}$ and a certain restriction on the m x n loss matrix of the generic problem. Examples violating the loss matrix condition are shown to have rate no faster than $O(N^{-1/2})$.

Additional results are presented when $m = n = 2$ and P_{θ_α}, for $\alpha = 1,2,\ldots$, depends on a fixed, but unknown, nuisance parameter $\tau = (\tau_1,\ldots,\tau_s)$ in a non-empty open set of Euclidean s-space. Under suitable regularity conditions on the likelihood ratio of P_1 and P_0 at the point τ, an asymptotic convergence theorem, uniform in $\theta \in \Omega_\infty$ and of $O(N^{-(1/2)+\varepsilon})$, $\varepsilon > 0$, is proved for the regret risk function of the procedure obtained by substituting the estimate $\bar{h}$ for $p(\theta)$ and a suitably chosen unbiased estimate $\bar{k} = N^{-1} \sum_{\alpha=1}^{N} k(X_\alpha)$ for τ. Theorems which are jointly uniform in $\theta \in \Omega_\infty$ and $\tau \in C$, a compact subset of R^s, are also given. When $s = 1$, two theorems dropping the factor $N^{+\varepsilon}$ in the convergence rate are established under appropriate restrictions.

Many examples illustrating the extent, applicability, necessity, and non-vacuity of the various theorems are added for completeness.

The emphasis throughout the thesis is on obtaining optimal asymptotic procedures in the sense of uniform regret risk convergence.

INTRODUCTION

The idea of the compound decision problem was first presented by Robbins in [10]*. When a large number of decision problems of identical nature occur, then the compound approach is applicable. In his paper, Robbins gave an example illustrating that when there are a large number of testing problems between two normal distributions $N(-1,1)$ and $N(1,1)$, then there exists a compound procedure whose risk is uniformly close to the risk of the best "simple" procedure based on knowing the proportion of component problems in which $N(1,1)$ is the governing distribution. This compound procedure depended on data from all component problems. Also in [10], heuristic arguments were given to illustrate that such a phenomenon could be expected more generally.

Hannan in [5] (see also Hannan and Robbins [7]) extended this result of Robbins to two arbitrary fully specified distributions; while simultaneously strengthening the conclusion by replacing "simple" by "invariant." Furthermore, in [7] it is shown that when the number of component problems is large, the compound procedure given has risk which is ε-better than the available minimax procedure.

In this thesis, we improve and generalize some of the results of Hannan and Robbins. Specifically, we examine asymptotically the difference in the risks (the regret risk function) of certain compound procedures and the empirical Bayes "non-simple" procedures.

*Numbers in square brackets refer to the bibliography.

In Chapter I, the general finite compound decision problem is presented. Also, we define "simple" Bayes procedures, which in turn motivate a class of "non-simple" compound decision procedures based on estimates of the empirical distribution on the finite parameter space. Theorem 1 solves the necessary estimation problem, while Corollary 1 and Lemma 5 set the stage for later developments.

In Chapter II, we treat the case of compound testing between two completely specified distributions P_o and P_1. Theorem 2 extends the basic theorem of Hannan and Robbins ([7], Theorem 4) by strengthening the asymptotic convergence rate of the regret risk function. Two additional theorems (Theorems 3 and 4) are proved. Both of these theorems give faster convergence rates under certain continuity requirements on P_o and P_1.

In Chapter III, we extend the results of Chapter II where possible to the general finite compound decision problem of Chapter I. Theorem 5 generalizes Theorem 2. Counter-examples to generalizations of Theorems 3 and 4 are given. However, by restrictions on the loss matrix of the component problem, Theorem 6 presents a suitable extension of Theorem 4.

In Chapter IV, the compound testing problem between two distributions in the presence of a nuisance parameter is considered. Convergence theorems for the regret risk function are given under suitable regularity conditions in the nuisance parameter.

At this point we introduce notation which will be used consistently throughout this thesis.

Let R^m be m-dimensional Euclidean space (R^1 will be denoted simply by R). Let $x = (x_o,\ldots,x_{m-1})$ and $y = (y_o,\ldots,y_{m-1})$ be vectors in R^m. Define the vector $xy = (x_o y_o,\ldots,x_{m-1}y_{m-1})$. The inner product and norm of R^m will be denoted respectively by $(x,y) = \sum_{i=o}^{m-1} x_i y_i$ and $\|x\| = (x,x)^{1/2}$. The inner product $(.,.)$ and norm $\|.\|$ notations will refer exclusively to R^m unless otherwise noted. Also, we will use $|x|$ to denote $\max_i |x_i|$.

Operator notation will be used to indicate integration. Let $(S,\mathcal{F},P)$ be any finite measure space with $\mathcal{F}$ a σ-field on S and P a finite measure on $(S,\mathcal{F})$. If X(s) is any real-valued integrable function on S, then PX will be used to denote the integral $\int X(s)dP(s)$. If P is a probability measure and X is a real-valued random variable, then PX denotes the expected value of X.

Also, we will make extensive use of the following notation for the characteristic function of a set A. The characteristic function of A will be denoted simply by A enclosed in square brackets; that is,

$$[A](a) = \begin{cases} 1 & \text{if } a \in A. \\ 0 & \text{if } a \notin A. \end{cases}$$

In reference to the previous paragraph, if F is a set of $\mathcal{F}$ and X(s) is any real-valued integrable function, then the P measure of F is given by P[F] and the definite integral $\int_F X(s)dP(s)$ by P(X[F]).

We will adopt the notation of Halmos ([4], Chapter VIII) to indicate induced measures under measurable transformations. Let T be a measurable transformation from $(S,\mathcal{F},P)$ into $(S',\mathcal{F}')$, where $\mathcal{F}'$ is a σ-field on S'. Then, let PT^{-1} denote the finite measure induced on $(S',\mathcal{F}')$ under the transformation T. The measure PT^{-1} is defined by the identity $PT^{-1}[F'] = P[T^{-1}(F')]$ for all $F' \in \mathcal{F}'$.

Finally, we shall make repeated use of the Berry-Esseen normal approximation theorem (see Loève [9], p. 288). This theorem, for simplicity, will be referred to by the letters B-E and the uniform constant in the bound by β. The standard normal distribution function will be denoted by $\Phi(\cdot)$ and the standard normal density by $\Phi'(\cdot)$.

Further notation will be introduced as needed.

CHAPTER I

THE FINITE COMPOUND DECISION PROBLEM

1. Statement of the Problem.

Consider the following finite statistical decision problem. Let U be a random variable (of arbitrary dimensionality) known to have one of m possible distributions P_θ, θ in the finite parameter space $\Omega = \{0,\ldots,m-1\}$. Based on observing U we are required to make a decision $d \in \mathcal{D} = \{0,\ldots,n-1\}$ incurring loss $L(i,j)$ (or L_i^j) if '$d = j$' when U is distributed as P_i, $i = 0,\ldots,m-1$; $j = 1,\ldots,n-1$.

If we simultaneously consider N decision problems each with this generic structure, then the N-fold global problem is called a finite compound decision problem. More precisely, let X_α, $\alpha = 1,\ldots,N$ be N independent observations each distributed as P_{θ_α} with θ_α ranging in Ω. Based on all N observations, a decision d_α in $\mathcal{D}$ is to be made for each of the N component problems. For the α^{th} subproblem, the decision '$d_\alpha = j$' represents selecting the j^{th} column of the m x n loss matrix. Note that in the case here considered all N decisions are held in abeyance until all random variables X_α, $\alpha = 1,\ldots,N$ have been observed.

In considering compound problems of the type described above, most of the results are of an asymptotic nature; that is, as $N \to \infty$. Hence, it will be convenient to adopt the following viewpoint. Let Ω_∞ be the set of all sequences $\theta = \{\theta_\alpha | \alpha = 1,2,\ldots\}$ where θ_α ranges in Ω. Consider now the above-stated compound problem (for N finite) as imbedded in the denumerable compound decision problem indexed by $\theta \in \Omega_\infty$, $\theta = \{\theta_\alpha\}$. Let P_θ be the product probability measure $\times_{\alpha=1}^{\infty} P_{\theta_\alpha}$. The above N-stage compound problem is equivalent to the compound problem obtained by observing the first N members of the sequence of random variables $\{X_1, X_2, \ldots\}$ distributed as P_θ, $\theta \in \Omega_\infty$.

Before proceeding, we introduce the following notation. With U as the generic name for the random variables X_α of the component problems, assume there exists a σ-finite measure μ dominating $\{P_o, \ldots, P_{m-1}\}$ such that the measurable densities

$$(1) \qquad f_i(u) = \frac{dP_i}{d\mu}(u) \leq K \qquad \text{a.e. } \mu$$

for some $K < \infty$. There is no loss of generality in this assumption since we may always choose $\mu = \sum_{i=o}^{m-1} P_i$ and $K = 1$.

Also in referring to the m x n matrix of losses $L(i,j)$ or L_i^j, the rows will be denoted by L_i, the columns by L^j, and the difference $L(i,k) - L(i,j)$ by L_i^{kj}, $i = 0,\ldots,m-1$; $j,k = 0,\ldots,n-1$.

2. Decision Procedures.

For the compound decision problem, a decision procedure may depend on the full observation $X = (X_1,\ldots,X_N)$. Any N x n matrix of measurable functions $T(x) = (t_{\alpha j}(x))$ will be called a randomized decision function (procedure) for the compound decision problem if for $\alpha = 1,\ldots,N$; $j = 0,\ldots,n-1$, $t_{\alpha j}(x) = \Pr\{d_\alpha = j | x\}$ and $\sum_{j=0}^{n-1} t_{\alpha j}(x) = 1$. The α^{th} row of $T(x)$ will be denoted by $t^{(\alpha)}(x) = (t_{\alpha o}(x),\ldots,t_{\alpha n-1}(x))$.

The decision function $T(x)$ is said to be simple if there exist functions $t_j(.)$, $j = 0,\ldots,n-1$ such that $t^{(\alpha)}(x) = (t_o(x_\alpha),\ldots,t_{n-1}(x_\alpha))$ for $\alpha = 1,\ldots,N$. A simple decision function will be denoted by $t = (t_o,\ldots,t_{n-1})$.

With N fixed and $\theta \in \Omega_\infty$ we denote by $R(\theta,T)$ the risk function for the compound decision procedure $T(x)$. This risk is defined to be the average of the component risks $R_\alpha(\theta,T) = P_\theta(L_{\theta_\alpha}, t^{(\alpha)}(X))$, for each subproblem, $\alpha = 1,\ldots,N$. Hence

$$(2) \qquad R(\theta,T) = N^{-1} \sum_{\alpha=1}^{N} R_\alpha(\theta,T) = P_\theta W(\theta,T(X)),$$

where $W(\theta,T(x)) = N^{-1} \sum_{\alpha=1}^{N} (L_{\theta_\alpha}, t^{(\alpha)}(x))$.

The risk (2) may be considerably simplified in the case of a simple decision function. For the sequence $\theta \in \Omega_\infty$ and $i = 0,\ldots,m-1$, define the relative frequencies, $p_i(\theta) = N^{-1} \sum_{\alpha=1}^{N} [\theta_\alpha = i]$, of problems in the first N problems in which the distribution P_i governs. The vector $p(\theta)$ will be called the empirical distribution on Ω.

Let $t = (t_o,\ldots,t_{n-1})$ be a simple decision function. The loss incurred in using procedure t is

$$(3) \qquad \begin{aligned} W(\theta,t) &= N^{-1} \sum_{\alpha=1}^{N} (L_{\theta_\alpha}, t(x_\alpha)) \\ &= \sum_{i=o}^{m-1} p_i(\theta) \{AV_{\theta_\alpha = i} (L_{\theta_\alpha}, t(x_\alpha))\} \end{aligned} \quad ,$$

where $AV_{\theta_\alpha = i}$ indicates the numerical average on the Np_i values $\theta_\alpha = i$. Now since $(L_{\theta_\alpha}, t(x_\alpha))$ for $\theta_\alpha = i$ are independent identically distributed random variables with mean $\rho_i(t) = P_i(L_i, t(U))$, we may express their expected average as $\rho_i(t)$ to obtain from (2) and (3),

$$(4) \qquad R(\theta,t) = \sum_{i=o}^{m-1} p_i(\theta)\, \rho_i(t) = (p(\theta), \rho(t)).$$

Let $\xi = (\xi_o,\ldots,\xi_{m-1})$ be any vector in m-dimensional Euclidean space. Let $t_j(u) \geq 0$, $j = 0,\ldots,n-1$ be a set of measurable functions such that $\sum_{j=0}^{n-1} t_j(u) = 1$. Define the function $\psi(\xi,t)$ as follows:

$$(5) \qquad \psi(\xi,t) = (\xi,\rho(t)).$$

Note that for $\xi = p(\theta)$ the function ψ becomes the risk function (4) for the simple decision procedure t.

The problem of choosing $t(u)$ to minimize $\psi(\xi,t)$ for fixed ξ is straightforward. From (1) and (5), we have

$$(6) \qquad \psi(\xi,t) = \mu \sum_{j=o}^{n-1} (\xi,L^j f(U))\, t_j(U) \;.$$

Therefore, (6) is minimized in t for fixed ξ by any vector function t_ξ (defined a.e. μ) which is chosen as a probability distribution concentrating on the columns L^j minimizing the quantities $(\xi,L^j f(u))$. That is, t_ξ is of the form

$$(7) \qquad t_{\xi,j}(u) = 1,\ 0 \text{ or arbitrary, for } (\xi,L^j f(u)) \quad <,\ >,\ \text{or} = \min_{\nu\neq j}(\xi,L^\nu f(u)),$$

such that $t_{\xi,j}(u) \geq 0$ for $j = 0,\ldots,n-1$ and $\sum_{j=o}^{n-1} t_{\xi,j}(u) = 1$ a.e. μ .

Note that if ξ is a bona fide a priori distribution, $(0 \leq \xi_i,\ \sum_{i=o}^{m-1} \xi_i = 1)$, then such a t_ξ would be a decision procedure Bayes against ξ.

We observe that any randomized procedure of the form (7) minimizing $\psi(\xi,t)$ may be replaced by a non-randomized version which also minimizes $\psi(\xi,t)$ for fixed ξ. In particular, one such non-randomized version is given by the coordinate functions

$$(8) \qquad t'_{\xi,j}(u) = \begin{cases} 1 & \text{if } (\xi, L^j f(u)) < \text{ or } \leqq (\xi, L^k f(u)) \\ & \text{according as } k < j \text{ or } k > j \\ 0 & \text{otherwise.} \end{cases}$$

To see that (8) is of the form (7) we merely note that $t'_\xi(u) = (t'_{\xi,o}(u), \ldots, t'_{\xi,m-1}(u))$ is a probability distribution concentrating on the first column minimizing the quantities $(\xi, L^j f(u))$. In what follows we restrict ourselves to the non-randomized version t'_ξ of the Bayes procedure t_ξ.

In [6], p. 102, Hannan has given a useful inequality for Bayes rules. A statement and proof of a similar result is given here.

<u>Lemma 1.</u>

Let X be a space closed under subtraction. Let $M(x,y)$ be a real-valued function on $X \times Y$ such that $M(\cdot,y)$ is linear on X for each $y \in Y$ and $\inf_y M(x,y)$ is attained for each $x \in X$. Define $f(x) = \inf_y M(x,y)$ and let $y(x)$ be any Y-valued function such that $f(x) = M(x,y(x))$ on X. Then, if $x, x' \in X$,

$$0 \leqq M(x,y(x')) - f(x) \leqq M(x-x',y(x')) - M(x-x',y(x)).$$

Proof. The lower inequality results from the definition of $f(x)$ and the upper inequality follows by adding the non-negative term $M(x',y(x)) - f(x')$.

Now define for $\xi \in R^m$ the function

$$(9) \qquad \phi(\xi) = \inf_t \psi(\xi,t) = (\xi, \rho(t_\xi)).$$

The last equality in (9) follows by noting that (7) minimizes $\psi(\xi,t)$. Observing that $(\xi,\rho(t))$ is linear in ξ and ρ, Lemma 1 and (9) yield

Corollary 1.

If $\xi, \xi' \in R^m$, then

(10) $$0 \leqq \psi(\xi,t_{\xi'}) - \phi(\xi) \leqq (\xi-\xi',\rho(t_{\xi}') - \rho(t_{\xi})).$$

This corollary inspires the non-simple rule to be adopted later (see (12)). If $p' \in R^m$ is a good approximation to $p(\theta)$ in the sense that $\|p'-p(\theta)\|$ is small, then Corollary 1 says that the simple procedure $t_{p'}(u)$ has risk within $\|p'-p(\theta)\| \, \|\rho(t_{p'}) - \rho(t_{p(\theta)})\|$ of the minimum attainable risk in the class of all simple procedures, given by $\phi(p(\theta))$. Therefore, not knowing $p(\theta)$ in general, we seek estimates $\hat{p} = \hat{p}(x_1,\ldots,x_N)$ of $p(\theta)$ which with the aid of Lemma 5 take advantage of the risk approximation of Corollary 1.

3. Estimation of Empirical Distributions on Ω.

The results in this section are based on some unpublished lecture notes of Hannan [8].

Let $\mathcal{P}$ be the class of all distributions on $\Omega = \{0,\ldots,m-1\}$; that is, $\mathcal{P} = \{\eta | \eta \in R^m, \eta_i \geqq 0, \sum_{i=0}^{m-1} \eta_i = 1\}$. For $\eta \in \mathcal{P}$ define the probability mixture $P_\eta = \sum_{i=0}^{m-1} \eta_i P_i$ with μ-density $f_\eta(u) = (\eta,f(u))$. The class of all distributions $\mathcal{P}$ is said to be identifiable if for any $\eta, \eta' \in \mathcal{P}$, $f_\eta(u) = f_{\eta'}(u)$ a.e. μ implies that $\eta = \eta'$.

Let $L_1(\mu)$ and $L_2(\mu)$ be the function spaces of μ-integrable and μ-square integrable functions respectively. The usual norm and inner product for $f, g \in L_2(\mu)$ will be denoted respectively by $\|f\|_\mu$ and $(f,g)_\mu$.

Lemma 2.

The class $\mathcal{P}$ is identifiable if and only if the set of densities $\{f_o,\ldots,f_{m-1}\}$ are linearly independent in $L_1(\mu)$.

Proof. Sufficiency. Let $f_\eta(u) = f_{\eta'}(u)$ a.e.μ. Then, $(\eta-\eta', f(u)) = 0$ a.e.μ and by linear independence of $\{f_o,\ldots,f_{m-1}\}$ it follows that $\eta_i = \eta_i'$ for $i = 0,\ldots,m-1$. Hence, $\eta = \eta'$ and $\mathcal{P}$ is identifiable.

Necessity. Let $\mathcal{P}$ be identifiable and let $c \in R^m$ be such that $(c,f(u)) = 0$ a.e.μ. Define c_i^+ and c_i^- as the positive and negative parts of c_i. Then $0 = \mu(c,f(u)) = \sum_{i=o}^{m-1} c_i$ and hence $\sum_{i=o}^{m-1} c_i^+ = \sum_{i=o}^{m-1} c_i^-$. If $\sum_{i=o}^{m-1} c_i^+ > 0$, define $d_i^+ = (\sum_{i=o}^{m-1} c_i^+)^{-1} c_i^+$ and $d_i^- = (\sum_{i=o}^{m-1} c_i^+)^{-1} c_i^-$. Then, $f_{d^+}(u) = f_{d^-}(u)$ a.e.μ and by identifiability of $\mathcal{P}$, $d_i^+ = d_i^-$ for all i. Hence, $c_i = c_i^+ - c_i^- = 0$ for all i and $c = 0$. Thus, necessity is proved.

A vector function $h = (h_o,\ldots,h_{m-1})$ with coordinate functions $h_i \in L_1(\mu)$ is an unbiased estimate for the class $\mathcal{P}$ if $P_\eta h = \eta$ for all $\eta \in \mathcal{P}$. Under the condition of identifiability of the class $\mathcal{P}$, existence of unbiased estimates for $\mathcal{P}$ will be shown. Henceforth, in accord with Lemma 2, the set of densities $\{f_o,\ldots,f_{m-1}\}$ are assumed to be linearly independent in $L_1(\mu)$. Let $\mathcal{E}$ be the class of all unbiased estimates for the class $\mathcal{P}$.

Lemma 3.

A necessary and sufficient condition for $h \in \mathcal{E}$ is that $P_i h = \varepsilon_i = (\delta_{io}, \ldots, \delta_{i\, m-1})$ for $i = 0, \ldots, m-1$, where δ_{ij} is the Kronecker δ.

Proof. Sufficiency. If $(P_i h_j)$ is the identity matrix, then $P_\eta h = \eta(P_i h_j) = \eta$ for all $\eta \in \mathcal{P}$.

Necessity. Observe that $\varepsilon_i \in \mathcal{P}$ and unbiasedness of h imply $P_{\varepsilon_i} h = \varepsilon_i$; that is $P_i h = \varepsilon_i$.

The following subclass of $\mathcal{E}$ is of particular interest. Let $\mathcal{H}$ be the subclass of $\mathcal{E}$ such that if $h \in \mathcal{H}$, $h_j \in L_2(\mu)$ for $j = 0, \ldots, m-1$, where $h = (h_o, \ldots, h_{m-1})$.

Let S be any subspace of $L_2(\mu)$ and $S^\perp$ be the orthogonal complement of S in $L_2(\mu)$. For any $g \in L_2(\mu)$, denote by g_S, $g_{S^\perp}$ the projection of g on S and $S^\perp$ respectively. Note that if $g \in L_2(\mu)$, $g = g_S + g_{S^\perp}$.

We now give a theorem which proves the existence of unbiased estimates for $\mathcal{P}$ and which yields the structure of the class $\mathcal{H}$. For $j = 0, \ldots, m-1$, let S_j be the subspace of $L_2(\mu)$ spanned by $\{f_i \mid i \neq j\}$. Let S be the subspace of $L_2(\mu)$ spanned by $\{f_o, \ldots, f_{m-1}\}$.

Theorem 1.

The class $\mathcal{H}$ is non-empty. Furthermore, $h \in \mathcal{H}$ if and only if $h(u) = f^*(u) + g(u)$ a.e. μ, where $f_j^*(u) = (f_{jS_j^\perp}(u))\,(\|f_{jS_j^\perp}\|_\mu^2)^{-1}$ and $g_j(u) \in S^\perp$ for $j = 0, \ldots, m-1$.

Proof. Note that since $f_j^* \varepsilon S_j^{\perp}$, $P_i f_j^* = (f_j^*, f_i)_\mu = 0$ for all $i \neq j$. Also, we have that $P_i f_i^* = (f_i^*, f_i)_\mu = 1$. Thus, by Lemma 3, $f_i^* \varepsilon \mathcal{E}$ (and hence $\mathcal{H}$ is non-empty since $f_i^* \varepsilon L_2(\mu)$). Sufficiency follows by observing that $P_i g_j = (g_j, f_i)_\mu = 0$ for $i,j = 0,\ldots,m-1$.

Conversely, if $h \varepsilon \mathcal{H}$, let $h = h_S + h_{S^\perp}$ having coordinate functions $h_j = h_{jS} + h_{jS^\perp}$ for $j = 0,\ldots,m-1$. Since $h_{jS^\perp}$ is in the orthogonal complement of S for $j = 0,\ldots,m-1$, $h_S \varepsilon \mathcal{H}$. Hence, $(f_j^* - h_{jS}, f_i)_\mu = 0$ for $i,j = 0,\ldots,m-1$. But this implies $f_j^* - h_{jS}$ is in $S^\perp$ as well as S. Hence, $f^* = h_S$ a.e. μ. Necessity follows by defining $g = h_{S^\perp}$.

Observe that the functions f_i^* of Theorem 1 form the dual basis to $\{f_o,\ldots,f_{m-1}\}$ in the conjugate space of the subspace S.

Corollary 2.

There exist $h \varepsilon \mathcal{E}$ such that $|h_i(u)| \leq M$ a.e. μ for $i = 0,\ldots,m-1$ and M finite.

Proof. Choose $h_i(u) = f_i^*(u)$ for $i = 0,\ldots,m-1$. Then, since the f_i^*'s lie in S, they are essentially bounded as linear combinations of the essentially bounded densities $\{f_o,\ldots,f_{m-1}\}$.

The importance of the class $\mathcal{E}$ in obtaining estimates for $p(\theta)$ can now be seen. Let $X = (X_1,\ldots,X_N)$ be the random observation for the N-fold compound problem stated earlier. Define by use of the kernel function $h \varepsilon \mathcal{E}$ the random variable

(11) $$\overline{h}(X) = N^{-1} \sum_{\alpha=1}^{N} h(X_\alpha) \quad .$$

This equation yields an unbiased estimate of the empirical distribution $p(\theta)$ for all $\theta \in \Omega_\infty$, since $P_\theta \bar{h}(X) = N^{-1} \sum_{\alpha=1}^{N} \varepsilon_{\theta_\alpha} = p(\theta)$. If $h \in \mathcal{E}$ and h is bounded as in Corollary 2, then $\bar{h}(x)$ inherits this boundedness through (11).

Consider now the subclass $\mathcal{H}$ of $\mathcal{E}$. If $h = (h_o, \ldots, h_{m-1}) \in \mathcal{H}$, then boundedness of the densities f_i implies $P_i h_j^2(U) < \infty$. Denote the variance of h_j under P_i for $i,j = 0,\ldots,m-1$ as $\sigma_i^2(h_j)$.

Lemma 4.

If $h \in \mathcal{H}$, then $P_\theta \|\bar{h}-p(\theta)\|^2 \leqq C^2 N^{-1}$, where $C^2 = \max_i \sum_{j=o}^{m-1} \sigma_i^2(h_j)$.

Proof. By direct computation, we have

$$P_\theta \|\bar{h}-p(\theta)\|^2 = \sum_{j=o}^{m-1} P_\theta(\bar{h}_j - p_j(\theta))^2$$

$$= N^{-1} \sum_{j=o}^{m-1} \sum_{i=o}^{m-1} p_i(\theta)\, \sigma_i^2(h_j)$$

$$\leqq C^2 N^{-1}.$$

4. Non-simple Decision Functions.

With $h \in \mathcal{H}$ and the estimate $\bar{h}(X)$ of $p(\theta)$ given by (11), we now define a non-simple decision function which results from substituting $\bar{h}(X)$ for $p(\theta)$ in $t_{p(\theta)}$ as given by (7) (see Hannan and Robbins [7], p. 44). In so doing, we shall confine ourselves to that particular non-randomized version of $t_{p(\theta)}$ given by (8) and denoted by $t'_{p(\theta)}$. The resulting non-simple, non-randomized decision procedure consists of the N vector functions $t'_{\bar{h}}(x_\alpha) = (t'_{\bar{h},o}(x_\alpha),\ldots,$ $t'_{\bar{h},n-1}(x_\alpha))$ for $\alpha = 1,\ldots,N$, where

$$(12)\qquad t'_{\bar{h},j}(x_\alpha) = \begin{cases} 1 \text{ if } (\bar{h}, L^j f(x_\alpha)) < \text{ or } \leqq (\bar{h}, L^\nu f(x_\alpha)) \\ \quad \text{according as } \nu < j \text{ or } \nu > j \\ 0 \text{ otherwise,} \end{cases}$$

$j = 0,\ldots,n-1$.

The question immediately arises regarding optimality properties of the procedure $t'_{\bar{h}}$. As a partial answer to this question, consider the function

$$(13)\qquad R(\theta,T) - \phi(p(\theta))$$

for the decision function $T(x)$ and $\theta \in \Omega_\infty$. This function will be called the <u>regret risk function</u> against simple decision functions for the decision procedure $T(x)$. A worthy defensive goal is to select a decision procedure $T(x)$ which makes the regret risk function small uniformly in $\theta \in \Omega_\infty$. In Chapters II, III, and IV it will be shown that the procedure $t'_{\bar{h}}$ (or a slightly modified version thereof), has, under suitable conditions, good asymptotic properties in the sense that its regret risk function given by (13) is close to zero uniformly in $\theta \in \Omega_\infty$ for N large.

We now give a useful decomposition lemma for the risk $R(\theta,T)$ in (13) for $T(x)$ such that

$$(14)\qquad T^{(\alpha)}(x) = t_\zeta(x_\alpha) \quad ,$$

where $\zeta = \zeta(x) = \zeta(x_1,\ldots,x_N)$ takes its values on a finite Euclidean space R^k and $t_\zeta(u) = (t_{\zeta,o}(u),\ldots,t_{\zeta,n-1}(u))$ is defined on $R^k \times U_{i=o}^{m-1} S_i$ with $S_i = \{u|f_i(u) > 0\}$ such that $\sum_{j=o}^{n-1} t_{\zeta,j}(u) = 1$, $t_{\zeta,j}(u) \geqq 0$.

Lemma 5.

Let $T(x)$ be a compound decision function of the form (14) and let $\theta \varepsilon \Omega_\infty$. Then,

$$(15) \qquad R(\theta,T) = P_\theta(p(\theta),\ \rho(t_\zeta))$$

$$+ N^{-1} \sum_{\alpha=1}^{N} \ \sum_{k\neq j} P_\theta P_{\theta_\alpha} L_{\theta_\alpha}^{kj} \ t_{\zeta(\alpha),k}(U)\ t_{\zeta,j}(U) \quad ,$$

where $\rho_i(t_\zeta) = P_i(L_i,t_\zeta(U))$ and $\zeta^{(\alpha)} = \zeta(x_1,\ldots,x_{\alpha-1},u,x_{\alpha+1},\ldots,x_N)$ and the P_{θ_α} integral in each of the N-terms of the second term of (15) is on U.

Proof. Fix $\alpha = 1,\ldots,N$ and express $P_\theta(L_{\theta_\alpha},\ T^{(\alpha)}(X))$ as an iterated integral, make a change of variable, and perform an added integration as follows,

$$(16) \qquad P_\theta(L_{\theta_\alpha},T^{(\alpha)}(X)) = \int(L_{\theta_\alpha},t_{\zeta(x)}(x_\alpha))\ dP_{\theta_\alpha}(x_\alpha)\Pi_{i\neq\alpha}dP_{\theta_i}$$

$$= \int(L_{\theta_\alpha},t_{\zeta^{(\alpha)}}(u))\ dP_{\theta_\alpha}(u)\Pi_{i\neq\alpha}dP_{\theta_i}$$

$$= \int(L_{\theta_\alpha},t_{\zeta^{(\alpha)}}(u))\ dP_{\theta_\alpha}(u)\Pi_i\ dP_{\theta_i}$$

$$= P_\theta P_{\theta_\alpha}(L_{\theta_\alpha},t_{\zeta^{(\alpha)}}(u)) \ ;$$

where $P_\theta P_{\theta_\alpha}$ represents an iterated integral. Writing $t_{\zeta^{(\alpha)}}(u) = t_{\zeta^{(\alpha)}}(u) - t_\zeta(u) + t_\zeta(u)$ in the right-hand side of (16) and averaging over all α, we have

$$(17) \qquad R(\theta,T) = N^{-1} \sum_{\alpha=1}^{N} P_\theta P_{\theta_\alpha}(L_{\theta_\alpha}, t_\zeta(U))$$

$$+ N^{-1} \sum_{\alpha=1}^{N} P_\theta P_{\theta_\alpha}(L_{\theta_\alpha}, t_{\zeta^{(\alpha)}}(U) - t_\zeta(U)) \quad .$$

The first term on the right-hand side of (17) may be simplified to $P_\theta(p(\theta),\rho(t_\zeta))$ by noting that for $\theta_\alpha = i, P_{\theta_\alpha}(L_{\theta_\alpha}, t_\zeta(U))$ are pointwise equal to $\rho_i(t_{\zeta(x)})$.

The second term in (17) may be simplified to the second term in (15) by observing that $(L_{\theta_\alpha}, t_{\zeta^{(\alpha)}}(u) - t_\zeta(u))$ is the difference of two inner products and that the components of $t_{\zeta^{(\alpha)}}(u)$ and of $t_\zeta(u)$ sum to unity.

CHAPTER II

ASYMPTOTIC RESULTS FOR THE COMPOUND TESTING PROBLEM FOR TWO COMPLETELY SPECIFIED DISTRIBUTIONS

1. Introduction and Notation.

In this chapter we discuss the compound decision problem of testing between two specified distributions. Robbins [10] showed that in the case where the component decisions were between $N(-1,1)$ and $N(1,1)$ there exists a decision function whose regret risk function approaches 0, uniformly in $\theta \epsilon \Omega_\infty$, as the number of problems N becomes large. Hannan and Robbins [7] extended this result to the case where the component decisions were between any two completely specified distributions. More extensive discussions of these and related results are given in [5], [7], and [11].

We treat the case as given in [5] and [7]. Three uniform convergence theorems for the regret risk function against simple decision functions will be given. The first of these theorems (Theorem 2 below) is an improvement of Theorem 4 in [7]. The improvement is in the rate of convergence. Before proceeding to the theorems some notational simplifications for testing between two distributions P_o and P_1 are in order.

Let $m = n = 2$ and take $L(0,0) = L(1,1) = 0$, $a = L(1,0) > 0$, and $b = L(0,1) > 0$. Specify the dominating measure to be $\mu = aP_1 + bP_o$, and note that by (1.1),

(1) $\qquad af_1(u) + bf_o(u) = 1 \qquad\qquad$ a.e. μ.

Define now the measurable transformation into $[0,1]$ by

(2) $$Z(u) = bf_o(u)$$

with (1) implying that

(3) $$1 - Z(u) = af_1(u) \qquad \text{a.e. } \mu.$$

Let μZ^{-1} be the measure induced on $[0,1]$ by the transformation (2) and denote by $\mu Z^{-1}(z)$ the non-normed left-continuous distribution function corresponding to μZ^{-1}. Note that $\mu Z^{-1}(z)$ has total variance $a + b$ since $\mu Z^{-1}(0) = 0$ and $\mu Z^{-1}(1+) = a + b$.

Identifying $t_\alpha(x) = t_{\alpha 1}(x)$ of Chapter I we can express a compound procedure by the N functions $t_\alpha(x)$, $\alpha = 1,\ldots,N$, since specification of $t_{\alpha o}(x)$ is not necessary as $t_{\alpha o}(x) = 1 - t_{\alpha 1}(x)$. Also, we represent a simple decision function by the single function t such that $t_\alpha(x) = t(x_\alpha)$.

For any p real, define the vector $\xi = (1 - p,p)$ in 2-space. In accord with (1.5) define for the simple decision function t the function

(4) $$\psi(p,t) = b(1-p)\, P_o t(U) + apP_1\, (1-t(U)).$$

A simple decision function minimizing (4) for fixed p, as given by (1.7), can with the aid of (1) and (2) be written as,

(5) $$t_p(u) = 1,0, \text{ or } \delta_p \text{ as } Z(u) <,>, \text{ or } = p, \text{ where } 0 \leq \delta_p \leq 1.$$

The non-randomized version of (5) with $\delta_p = 0$, corresponding to (1.8), shall be denoted by t'_p.

Also, by defining $\overline{\theta} = p_1(\theta)$, we may simply express the Bayes risk, given by (1.9), against $(1-\overline{\theta},\overline{\theta})$ on $\Omega = \{0,1\}$ as

(6) $$\phi(\overline{\theta}) = \inf_t \psi(\overline{\theta},t) = \psi(\overline{\theta},t_{\overline{\theta}}) \quad .$$

The assumption that P_o and P_1 are distinct implies that f_o and f_1

are linearly independent in $L_1(\mu)$. Furthermore, the choice of μ implies f_o and f_1 are essentially bounded functions. Thus, by Theorem 1 there exists a scalar function $h \varepsilon L_2(\mu)$ such that $P_i h(U) = i$ for $i = 0,1$; that is, identify h with h_1 of Theorem 1 and regard $\mathcal{H}$ as a class of scalar functions (h_o being defined as 1-h). For such an $h \varepsilon \mathcal{H}$, define for $i = 0,1$,

$$(7) \qquad \sigma_i^2(h) = P_i(h(U) - i)^2, \quad \bar{\sigma}^2(h) = \max_{i=o,1} \{\sigma_i^2(h)\},$$

and for any $0 \leqq p \leqq 1$,

$$(8) \qquad \sigma_p^2(h) = p\sigma_1^2(h) + (1-p)\, \sigma_o^2(h) \qquad .$$

From (1.11) we now have the unbiased scalar estimate $\bar{h}(X) = N^{-1} \sum_{\alpha=1}^{N} h(X_\alpha)$ of $\bar{\theta}$ and from (1.12) the associated compound decision rule (here slightly modified at $Z(x_\alpha) = 0$ or 1) given by

$t_h^*(x) = (t_{\bar{h}}^*(x_1), \ldots, t_{\bar{h}}^*(x_N))$, where, for $\alpha = 1,\ldots,N$,

$$(9) \qquad t_{\bar{h}}^*(x_\alpha) = \begin{cases} 1 & \text{if } Z(x_\alpha) < \bar{h},\ Z(x_\alpha) \ \varepsilon \ (0,1) \quad \text{or} \quad Z(x_\alpha) = 0 \\ 0 & \text{if } Z(x_\alpha) \geqq \bar{h},\ Z(x_\alpha) \ \varepsilon \ (0,1) \quad \text{or} \quad Z(x_\alpha) = 1 \end{cases} \qquad .$$

Observe that if $\bar{h} \ \varepsilon \ [0,1]$, then (9) is a decision procedure Bayes against a priori $(1-\bar{h},\bar{h})$ in the component problem.

The justification for modifying (9) at the endpoints $Z(x_\alpha) = 0$ and $Z(x_\alpha) = 1$ will become apparent if one considers the risk function $R(\theta,t)$ for any decision procedure $t(x) = (t_1(x),\ldots,t_N(x))$. The component loss for the α^{th} subproblem using $t(x)$ is given by $a\theta_\alpha(1-t_\alpha(x)) + b(1-\theta_\alpha)t_\alpha(x)$. Hence, this risk, as the expection of the average of the N component losses, can, with the aid of (2) and (3), be expressed as

$$(10)\quad R(\theta,t) = N^{-1}P_\theta\sum_{\alpha=1}^{N}\int\{\theta_\alpha(1-t_\alpha(x))(1-Z(x_\alpha))+(1-\theta_\alpha)t_\alpha(x)Z(x_\alpha)\}d\mu(x_\alpha).$$

Now note that in (10) if $t_\alpha(x) \neq 1$ for $Z(x_\alpha) = 0$ or if $t_\alpha(x) \neq 0$ for $Z(x_\alpha) = 1$ we may always redefine $t_\alpha(x)$ at these endpoints to achieve a risk which is at least as small as (10) (and maybe actually smaller, in which case t would be inadmissible). To avoid such a possibility with decision procedure (9) we have made the appropriate modifications at the endpoints $Z(x_\alpha) = 0$ and $Z(x_\alpha) = 1$, for $\alpha = 1,\ldots,N$.

2. An Inequality for the Regret Risk Function.

We shall develop a useful inequality (see (13)) for the regret risk function. We have already defined the procedure $t^*_{\overline{h}}$ by (9) in such a way that there is no contribution to the α^{th} term of the risk $R(\theta,t^*_{\overline{h}})$ in (10) at the endpoints $Z(x_\alpha) = 0$ or 1 for $\alpha = 1,\ldots,N$. The risk $R(\theta,t_{\underline{\theta}})$ has this same property since $R(\theta,t_{\underline{\theta}}) = R(\theta,t^*_{\underline{\theta}})$. Therefore, for convenience in notation, we define the restrictions of the P_i measures to $Z^{-1}(0,1)$ as follows: $P'_i(B) = P_i(B \cap Z^{-1}(0,1))$ for any Borel set B, $i = 0,1$. Also, observe that $\mu' = aP'_1 + bP'_0$ is the restriction of μ to $Z^{-1}(0,1)$.

Consider now the application of Lemma 5 to bound from above $R(\theta,t^*_{\underline{h}}) - \phi(\overline{\theta})$. With $t_\zeta = t^*_{\underline{h}}$ in Lemma 5, we bound the second term in the right-hand side of (1.15) from above by dropping all terms with negative coefficients $L^{kj}_{\theta_\alpha}$ and express $t^*_{\underline{h}(\alpha)}$ and $t^*_{\underline{h}}$ in their characteristic function form to obtain,

$$(11) \qquad N^{-1}\sum_{\alpha=1}^{N}\sum_{k\neq j} P_{\theta} P_{\theta_\alpha} L^{kj}_{\theta_\alpha} t^{*}_{\overline{h}^{(\alpha)},k}(U)\, t^{*}_{\overline{h},j}(U)$$

$$\leqq N^{-1}\, a\sum_{\alpha\varepsilon I_1} P_\theta P_1' [\overline{h}^{(\alpha)} \leqq Z < \overline{h}]$$

$$+ N^{-1}\, b\sum_{\alpha\varepsilon I_o} P_\theta P_o' [\overline{h} \leqq Z < \overline{h}^{(\alpha)}] \qquad ,$$

where $I_i = \{\alpha | \theta_\alpha = i\}$ for $i = 0,1$.

The integrand in the first term on the right-hand side of (1.15) can be expressed as $\psi(\overline{\theta},t^*_{\overline{h}})$ and, since $t'_{\overline{\theta}}$ is Bayes against $(1-\overline{\theta},\overline{\theta}),\phi(\overline{\theta})$ by (6) equals $\psi(\overline{\theta},t'_{\overline{\theta}})$. Hence, definition (4), expression of t'_{θ} and $t^*_{\overline{h}}$ in characteristic function form, and the definition of μ' yield

$$(12) \qquad (p(\theta),\rho(t^*_{\overline{h}})) - \phi(\overline{\theta})$$

$$= \mu'\{(1-\overline{\theta})Z([Z<\overline{h}]-[\overline{Z<\theta}]) + \overline{\theta}(1-Z)([\overline{h}\leqq Z]-[\overline{\theta}\leqq Z])\}$$

$$= \mu'\{(Z-\overline{\theta})([\overline{\theta}\leqq Z<\overline{h}]-[\overline{h}\leqq Z<\overline{\theta}])\} \qquad ,$$

where the second equality follows by set algebra and algebraic cancellation.

Equations (11) and (12) combine to yield the following inequality for the regret risk function:

$$(13) \qquad R(\theta,t^*_{\overline{h}})-\phi(\overline{\theta})$$

$$\leqq P_\theta \mu'\{(Z-\overline{\theta})([\overline{\theta}\leqq Z<\overline{h}]-[\overline{h}\leqq Z<\overline{\theta}])\}$$

$$+ N^{-1}\, a\sum_{\alpha\varepsilon I_1} P_\theta P_1'[\overline{h}^{(\alpha)}\leqq Z<\overline{h}]$$

$$+ N^{-1}\, b\sum_{\alpha\varepsilon I_o} P_\theta P_o'[\overline{h}\leqq Z<\overline{h}^{(\alpha)}] \qquad ,$$

where $I_i = \{\alpha | \theta_\alpha = i\}$ for $i = 0,1$.

When applying inequality (13) the three terms on the right-hand side will be denoted by A_N, B_N, and C_N respectively.

3. A Convergence Theorem of $O(N^{-1/2})$.

Sufficient conditions for uniform convergence (in $\theta\epsilon\Omega_\infty$) of $O(N^{-1/2})$ for the regret risk function of the procedure $t^*_{\bar{h}}$ will be given. Before proceeding to the theorem we state the following inequality: If C be a non-negative real number and if $N^{-1} \leq p \leq 1$, then

$$(14) \qquad N^{1/2} p \min \{1,(Np-1)^{-1/2} C\} \leq (1+C^2)^{1/2} p^{1/2} \quad .$$

Verification of inequality (14) is straightforward: If $(Np-1) \geq C^2$, then $N^{1/2}p(Np-1)^{-1/2}C = Cp^{1/2}(1-(Np)^{-1})^{-1/2}$ $\leq p^{1/2}(1+C^2)^{1/2}$, and if $(Np-1) \leq C^2$, then $N^{1/2}p = (Np)^{1/2} p^{1/2}$ $\leq (1+C^2)^{1/2} p^{1/2}$.

Theorem 2.

If $h \epsilon \mathcal{E}$ and $\epsilon L_3(P_i)$ for $i = 0,1$, then $R(\theta,t^*_{\bar{h}}) - \phi(\bar{\theta}) = O(N^{-1/2})$ uniformly in $\theta \epsilon \Omega_\infty$.

Proof. In inequality (13) we show: (i) $A_N = O(N^{-1/2})$ uniformly in $\theta \epsilon \Omega_\infty$, and (ii) B_N and C_N are of $O(N^{-1/2})$ uniformly in $\theta \epsilon \Omega_\infty$.

(i) Since $\mu'\{(Z-\bar{\theta})([\bar{\theta}\leq Z<\bar{h}] - [\bar{h}\leq Z<\bar{\theta}])\} \leq |\bar{h}-\bar{\theta}|(a+b)$ a.e. P_θ, then $N^{1/2} A_N \leq (a+b) P_\theta(N^{1/2}|\bar{h}-\bar{\theta}|) \leq (a+b) \sigma_{\bar{\theta}}(h) \leq (a+b) \bar{\sigma}(h)$. Independence of $\theta \epsilon \Omega_\infty$ for the upper bound implies uniformity and (i) is proved.

(ii) In bounding the term $N^{1/2}B_N$, we can assume without loss of generality that I_1 is non-void and $\sigma_1^2(h)>0$. If $\sigma_1^2(h) = 0$, then $\bar{h}^{(\alpha)} = \bar{h} + N^{-1}(h(u)-h(x_\alpha)) = \bar{h}$ a.e. $P_\theta \times P_1'$ for all $\alpha \epsilon I_1$, and hence $[\bar{h}^{(\alpha)} \leq Z<\bar{h}] = 0$ a.e. $P_\theta \times P_1'$ for $\alpha \epsilon I_1$; that is, $B_N = 0$.

Fix $\alpha \in I_1$ and N and let $\sigma_1 = \sigma_1(h) > 0$. Define $S = \sum_{i\in I_1, i\neq\alpha} (h(x_i)-1)$, $\sigma^2 = \mathrm{Var}(S)$, and $T = N(Z-\bar{\theta}) + 1 - \sum_{i\in I_o} h(x_i)$.

Then,

$$(15) \qquad [\bar{h}^{(\alpha)} \leqq Z < \bar{h}] = [T-h(x_\alpha) < S \leqq T-h(u)] \quad .$$

Apply the B-E theorem conditionally on u, x_α, and x_i, $i \in I_o$, to the normalized sum $\sigma^{-1}S$ at the endpoints $\sigma^{-1}(T-h(x_\alpha))$ and $\sigma^{-1}(T-h(u))$ and bound the resulting absolute difference in normal d.f.'s by $\Phi'(0)|h(u)-h(x_\alpha)|\sigma^{-1}$. Noting that $\sigma^2 = (N\bar{\theta}-1)\sigma_1^2$, the result from (15) is

$$(16) \qquad P_\theta P_1' [\bar{h}^{(\alpha)} \leqq Z < \bar{h}]$$
$$\leqq \min\{1, (N\bar{\theta}-1)^{-1/2}(\Phi'(0)\sigma_1^{-1} P_1'P_{\theta_\alpha}|h(U)-h(X_\alpha)| + 2\beta a_1)\}$$

where $a_1 = \sigma_1^{-3} P_1|h(U)-1|^3$.

Weakening the bound in (16) by the Schwarz inequality applied to $P_1'P_{\theta_\alpha}|h(U)-h(X_\alpha)| \leqq \{P_1'P_{\theta_\alpha}(h(U)-h(X_\alpha))^2\}^{1/2} \leqq 2^{1/2}\sigma_1$, and summing (16) over all $\alpha \in I_1$, we have $B_N \leqq a\,\bar{\theta}\min\{1, (N\bar{\theta}-1)^{-1/2} b_1\}$, where $b_1 = 2^{1/2}\Phi'(0) + 2\beta a_1$. Inequality (14) now yields the desired result $N^{1/2} B_N \leqq a(1+b_1^2)^{1/2}$.

A similar argument shows that $N^{1/2}C_N \leqq b(1+b_o^2)^{1/2}$, where $b_o = 2^{1/2}\Phi'(0) + 2\beta a_o$ with $a_o = \sigma_o^{-3}P_o|h(U)|^3$. Finally, since b_o and b_1 do not depend on $\theta \in \Omega_\infty$, (ii) is proved. The theorem now follows by (i), (ii) and inequality (13).

At this point it is worthwhile to make a few remarks regarding the assumptions on h in Theorem 2. By the choice of μ it is evident that f_o and f_1 are essentially bounded and hence Corollary 2 guarantees the existence of an estimate h which is also essentially bounded. Thus, it seems unnecessary to weaken the assumptions on h in Theorem 2 to include h's whose third absolute moments are finite under P_o and P_1.

The importance of bounded h's is also illustrated by the constructive procedure given by Hannan and Robbins ([7], pp. 42-43) for obtaining a uniformly bounded kernel estimate h which is unbiased and minimizes, for fixed p, $0<p<1$, $\sigma_p^2(h)$ given by (8).

However, we present now an example which shows that the enlarged class furnishes an unbounded unbiased estimate $\bar{h}(X)$ of $\bar{\theta}$ for all $\theta \in \Omega_\infty$ which is easy to compute when compared to the estimate $\bar{f}^*(x)$, given by Theorem 1 and (1.11).

Example. Let $X_\alpha = (X_{\alpha 1},\ldots,X_{\alpha n})$ be the random variable for the α^{th} subproblem. For each $\alpha = 1,\ldots,N$, assume $X_{\alpha 1},\ldots,X_{\alpha n}$ are n independent identically distributed random variables having one of two distributions $G_i(\cdot)$ for $i = 0,1$. Let $G_i(\cdot)$ be a normal distribution function with mean ω_i and variance σ^2, for $i = 0,1$. Assume $\omega_1 > \omega_o$. Let P_o and P_1 denote the respective product measures G_o^n and G_1^n. Denote by $g_i(\cdot)$ and $p_i(\cdot)$ for $i = 0,1$ the Lebesque densities of G_i and P_i respectively. Then $p_i(u) = \Pi_{j=1}^n g_i(u_j)$ for $i = 0,1$ is the joint density of the n independent random variables $U_1,\ldots, U_n$.

Observe that $p_i(u)$, for $i = 0,1$ are bounded and we may apply Theorem 1 and Corollary 2 to obtain a bounded estimate $f^*(u) = f^*(u_1,\ldots,u_n)$. By Theorem 1, $f^*(u) = p_{1S_1^\perp}(u) \, \|p_{1S_1^\perp}\|^{-2}$ where $p_{1S_1^\perp}(u) = p_1(u)-(p_o,p_1)||p_o||^{-2} \, p_o(u)$. The L_2 norms and inner product in these expressions are with respect to n-dimensional Lebesque measure. Simple linear space algebra therefore yields

$$(17) \qquad f^*(u) = \frac{||p_o||^2 \, p_1(u) - (p_o,p_1) \, p_o(u)}{||p_o||^2 \, ||p_1||^2 - (p_o,p_1)^2} .$$

We now compute the norms and inner product in (17). For i = 0,1

(18) $$\|p_i\|^2 = (2\pi\sigma^2)^{-n}\,\Pi_{j=1}^{n}\int_{-\infty}^{\infty} \exp\{-\sigma^{-2}(u_j-\omega_i)^2\}\,du_j = (2\pi^{1/2}\,\sigma)^{-n} \quad ,$$

where the second equality follows from the transformations $v_j = 2^{1/2}\,\sigma^{-1}(u_j-\omega_i)$ for $j = 1,\ldots,n$. Also

$$(p_o,p_1) = (2\pi\sigma^2)^{-n}\,\Pi_{j=1}^{n}\int_{-\infty}^{\infty} \exp\{-(2\sigma^2)^{-1}[(u_j-\omega_o)^2+(u_j-\omega_1)^2]\}\,du_j$$

$$= (2\pi\sigma^2)^{-n}\,c^n\Pi_{j=1}^{n}\int_{-\infty}^{\infty} \exp\{-\sigma^{-2}[u_j - \tfrac{1}{2}(\omega_o+\omega_1)]^2\}\,du_j$$

where $c = \exp\{-(2\sigma)^{-2}(\omega_1-\omega_o)^2\}$. The second equality follows by completing the square in u_j in the exponent of the n integrands. Transforming the n integrands in this last expression by $v_j = 2^{1/2}\sigma^{-1}[u_j- \frac{1}{2}(\omega_o+\omega_1)]$ for $j=1,\ldots,n$ will then yield $(p_o,p_1) = (2\pi^{1/2}\sigma)^{-n}\,c^n$. This result together with (18), when substituted into (17), furnishes the unbiased estimate

(19) $$f^*(u) = (2\pi^{1/2}\sigma)^n\,(1-c^{2n})^{-1}\,(p_1(u)-c^np_o(u)) \quad .$$

With $X = (X_1,\ldots,X_N)$ and $X_\alpha = (X_{\alpha 1},\ldots,X_{\alpha n})$ for $\alpha = 1,\ldots,N$, (19) can be used as a kernel function in (1.11) to give the following unbiased estimate of $\overline{\theta}$,

(20) $$\overline{f}^*(X) = N^{-1}\sum_{\alpha=1}^{N} f^*(X_\alpha)$$

$$= (2\pi^{1/2}\sigma)^n\,(1-c^{2n})^{-1}\,N^{-1}\sum_{\alpha=1}^{N}(p_1(X_\alpha)-c^np_o(X_\alpha))$$

$$= c_oN^{-1}\sum_{\alpha=1}^{N}\{\exp(c_1\sum_{j=1}^{n}(X_{\alpha j}-\omega_1)^2)-c^n\,\exp(c_1\sum_{j=1}^{n}(X_{\alpha j}-\omega_o)^2)\} \quad ,$$

where $c_o = (2)^{n/2}(1-c^{2n})^{-1}$ and $c_1 = -(2\sigma^2)^{-1}$. From (20) it is evident that the unbiased estimate $\overline{f}^*(X)$ of $\overline{\theta}$ is not easy to compute.

However, consider the following unbounded estimate of $\overline{\theta}$. Let $\overline{X}_\alpha = n^{-1}\sum_{j=1}^{n}X_{\alpha j}$ and $\overline{X} = N^{-1}\sum_{\alpha=1}^{N}\overline{X}_\alpha$. Define $h(X_\alpha) = (\omega_1-\omega_o)^{-1}(\overline{X}_\alpha-\omega_o)$.

Then, we have $P_{\theta_\alpha} h(X_\alpha) = \theta_\alpha$. Therefore, $h(X_\alpha)$ is an unbiased estimate of $\theta_\alpha = 0$ or 1. Hence, in accord with (1.11),

$$\overline{h}(X) = N^{-1} \textstyle\sum_{\alpha=1}^{N} h(X_\alpha) = (\omega_1-\omega_o)^{-1} (\overline{X} - \omega_o) \qquad (21)$$

is an unbiased estimate of $\overline{\theta}$ for all $\theta \varepsilon \Omega_\infty$. The computational advantage of (21) over (20) is apparent, and this example serves to illustrate the usefulness of the weakened assumptions on h in Theorem 2.

The above example can be generalized to any two distributions P_o and P_1 for which there exists a function ζ with $P_i|\zeta(U)|^3 < \infty$ and $\omega_i = P_i\zeta(U)$ for $i = 0,1,\omega_o \neq \omega_1$. Define $h(u) = (\omega_1-\omega_o)^{-1} (\zeta(u)-\omega_o)$. Then h satisfies the conditions of Theorem 2 and can be used as the kernel in (1.11). This is the type of estimate suggested by Robbins in [10] where he uses $\frac{1}{2}(\overline{X} + 1)$, with $\overline{X} = N^{-1} \sum_{\alpha=1}^{N} X_\alpha$, as an unbiased estimate of $\overline{\theta}$ in the compound testing problem where the α^{th} component problem is testing $N(-1,1)$ against $N(1,1)$ based on one observation X_α.

In the next section this generality of estimates is not retained. The proofs of Theorems 3 and 4 utilize stronger properties of h. Theorem 4 requires essential boundedness, while Theorem 3 has strong moment assumptions on h.

4. Convergence Theorems of Higher Order.

Convergence rates faster than that in Theorem 2 are obtainable under successively stronger sufficient conditions. The following conditions on the continuity of the induced distributions P_iZ^{-1} for $i = 0,1$ are pertinent.

(I) Let the induced distributions P_iZ^{-1} be continuous functions on $(0,1)$ for $i = 0,1$.

It is an immediate consequence of (I) that $\mu'Z^{-1}$ is continuous (and hence uniform continuity) on $[0,1]$.

To see this, note that $\mu' Z^{-1}(z) = \mu[0<Z(U)<z,\ Z(U)<1]$ implies that $\mu' Z^{-1}(0+) = \inf_{z>0} \mu' Z^{-1}(z) = 0 = \mu' Z^{-1}(0)$ and $\mu' Z^{-1}(1+) = \inf_{z>1} \mu' Z^{-1}(z) = \mu' Z^{-1}(1)$. These results together with left-continuity of $\mu' Z^{-1}(z)$ imply $\mu' Z^{-1}$ is continuous on $[0,1]$.

(II) Let λ be Lebesque measure and $P_i Z^{-1}$ be absolutely continuous with respect to λ, for $i = 0,1$. Let there exist a $K' < \infty$ such that

$$\frac{dP_i Z^{-1}}{d\lambda}(z) \leqq K' \qquad \text{a.e. } \lambda. \tag{22}$$

It is an immediate consequence of (II) that

$$\frac{d\mu' Z^{-1}}{d\lambda}(z) \leqq (a+b)\, K' \qquad \text{a.e. } \lambda. \tag{23}$$

We now prove with the aid of inequality (13) the following two uniform convergence theorems for the regret risk function.

Theorem 3.

Let $h \in \mathcal{E}$ be such that $P_i|h(U)-i|^k \leqq 2^{-1}\sigma_i^2(h)k!\ q^{k-2}$; $k = 2,3,\ldots,i=0,1$, and some $q > 0$. Then, if (I) holds $R(\theta,t_h^*) - \phi(\bar\theta) = o\,(N^{-1/2})$ uniformly in $\theta \in \Omega_\infty$.

Proof. We show (i) $A_N = o(N^{-1/2})$ uniformly in $\theta \in \Omega_\infty$ and (ii) B_N and C_N are $o(N^{-1/2})$ uniformly in $\theta \in \Omega_\infty$.

(i) Let $\varepsilon > 0$ be given. Under assumption (I), $\mu' Z^{-1}(z)$ is uniformly continuous on $[0,1]$ (and hence on R). Therefore, there exists a $\delta = \delta(\varepsilon) > 0$ such that $\mu' Z^{-1}[[z_1,z_2)] \leqq 8^{-1/2}\ \varepsilon$ whenever $|z_2-z_1| < \delta$. Choose N_0 sufficiently large such that $N_0 \geqq 8(\delta\varepsilon)^{-2}\ \{(a+b)\bar\sigma\}^2$, where $\bar\sigma^2 = \bar\sigma^2(h)$. Let $E = \{|\bar h-\bar\theta| \geqq \delta\}$ and observe that by Tchebichev's inequality,

(24) $$P_\theta[E] \leq N^{-1}\,\delta^{-2}\,\sigma^2_{\bar\theta}(h)$$
$$\leq N^{-1}\,\delta^{-2}\,\bar\sigma^2 \quad .$$

Consider now the term $A^2_{1,N} = N\{P_\theta\mu'(Z-\bar\theta)[\bar\theta \leq Z < \bar h]\}^2$. Use of the pointwise inequality $(Z-\bar\theta)[\bar\theta \leq Z < \bar h] \leq |\bar h-\bar\theta|[\bar\theta \leq Z < \bar h]$ in $A^2_{1,N}$, followed by the Schwarz integral inequality yields the bound $A^2_{1,N} \leq \sigma^2_{\bar\theta}(h)\; P_\theta\{\mu'[\bar\theta \leq Z < \bar h]\}^2$. In the second factor of this bound, partition the space under the P_θ integral into E and its complement E^c, noting that on E^c, $\mu'[\bar\theta \leq Z < \bar h] = \mu' Z^{-1}[[\bar\theta,\bar h)] \leq 8^{-1/2}\,\varepsilon$, while on E, $\mu'[\bar\theta \leq Z < \bar h] \leq (a+b)$. Hence, $A^2_{1,N} \leq \sigma^2_{\bar\theta}(h)(8^{-1}\varepsilon^2 + (a+b)^2\,P_\theta[E])$. Inequality (24) and the choice of N_o yield for $N \geq N_o$, $A_{1,N} \leq \frac{1}{2}\,\bar\sigma\varepsilon$.

By a similar argument we obtain $A_{2,N}=N^{1/2}\{P_\theta\mu'(\bar\theta-Z)[\bar h \leq Z < \bar\theta]\} \leq \frac{1}{2}\,\bar\sigma\varepsilon$. Observing that $N^{1/2}A_N = A_{1,N}+A_{2,N}$, the previous two inequalities yield $N^{1/2}A_N \leq \bar\sigma\varepsilon$. Since ε is arbitrary, and since both $\bar\sigma$ and N_o are independent of $\theta \in \Omega_\infty$, (i) is proved.

(ii) Let $\varepsilon > 0$ be given. By uniform continuity of $P_1' Z^{-1}(z)$ on R, there exists a $\delta' = \delta'(\varepsilon) > 0$ such that $P_1' Z^{-1}[[z_1,z_2)] \leq \frac{1}{2}\,\varepsilon^2$ if $|z_2-z_1| \leq \delta'$. The proof for the term B_N relies upon properly bounding the two terms on the right-hand side of the expression

(25) $$B_N = N^{-1}\,a\sum\nolimits_{\alpha\varepsilon I_1} P_1'\{[F]\,P_\theta[\bar h^{(\alpha)} \leq Z < \bar h]\}$$
$$+\; N^{-1}\,a\sum\nolimits_{\alpha\varepsilon I_1} P_1'\{(1-[F])P_\theta[\bar h^{(\alpha)} \leq Z < \bar h]\} \quad ,$$

Where $F = \{|Z-\bar\theta| \leq \delta'\}$. The two terms on the right-hand side of (25) will be denoted by $B_{1,N}$ and $B_{2,N}$ respectively.

We first bound the $B_{1,N}$ term in (25) by a B-E approximation argument. As in the proof of Theorem 2, we assume without loss of generality that $\sigma_1^2 = \sigma_1^2(h) > 0$ and I_1 is non-void. By a B-E approximation conditionally on u, x_α, and x_i, $i\varepsilon I_o$ applied to α^{th} summand in $B_{1,N}$ we have by

(15) and (16),

(26) $$P_1'\{[F]P_\theta[\overline{h}^{(\alpha)} \leq Z < \overline{h}]\}$$
$$\leqq \min\{P_1'[F], (N\overline{\theta}-1)^{-1/2}(\Phi'(0)\sigma_1^{-1}P_1'P_{\theta_\alpha}|h(U)-h(X_\alpha)|[F]+2\beta a_1 P_1'[F])\}.$$

Weakening in (26) by the Schwarz integral inequality to obtain $P_1'P_{\theta_\alpha}|h(U)-h(X_\alpha)|[F] \leqq 2^{1/2}\sigma_1\{P_1'[F]\}^{1/2}$, observing that our choice of δ' implies that $P_1'[F] \leqq \varepsilon^2$, and summing (26) over all $\alpha \in I_1$, the definition of $B_{1,N}$ and inequality (14) yield

(27) $$N^{1/2}B_{1,N} \leqq a\varepsilon^2 N^{1/2}\,\overline{\theta}\,\min\{1, (N\overline{\theta}-1)^{-1/2}(2^{1/2}\Phi'(0)\varepsilon^{-1} + 2\beta a_1)\}$$
$$\leqq a\varepsilon(\varepsilon+2^{1/2}\Phi'(0)+2\beta a_1\varepsilon).$$

Since ε is arbitrary and the bound in (27) is independent of $\theta \in \Omega_\infty$, we have

(28) $$\lim_{N\to\infty} N^{1/2}B_{1,N} = 0, \text{ uniformly in } \theta \in \Omega_\infty.$$

We now bound $B_{2,N}$ in (25) by Bernstein's exponential inequality given in the following theorem (see [2] for proof).

Theorem: (Bernstein).

Let $Y_1, Y_2, \ldots$ be a sequence of independent random variables with $\sigma_i^2 = \mathrm{Var}(Y_i)$ and such that $P|Y_i - PY_i|^k \leqq 2^{-1}\sigma_i^2 k!\,q^{k-2}$, for $k = 2,3,\ldots$; $i = 1,2,\ldots$, and some $q > 0$. Let $S_n = \sum_{i=1}^n (Y_i - PY_i)$ and $s_n^2 = \sum_{i=1}^n \sigma_i^2$. Then, for any $t > 0$, $P[|S_n| > ts_n] < 2\exp\{-(2+2qts_n^{-1})^{-1}t^2\}$.

Before using this theorem for bounding $B_{2,N}$ observe the following set inclusion,

$$\{|Z-\overline{\theta}| > \delta', \overline{h}^{(\alpha)} < Z < \overline{h}\}$$
$$\subset \{\overline{h}-\overline{\theta}>\delta'\} \cup \{\overline{h}^{(\alpha)}-\overline{\theta}<-\delta'\} \qquad .$$

Substituting this set inclusion into $B_{2,N}$ and observing that a simple change of variable implies $P_\theta P_1'[\overline{h}^{(\alpha)}-\overline{\theta}<-\delta'] \leqq P_\theta[\overline{h}-\overline{\theta}<-\delta']$ for all $\alpha \in I_1$, we obtain $B_{2,N} \leqq a\bar{\theta}P_{\theta}[|\bar{h}-\bar{\theta}|>\delta']$. Application of Bernstein's inequality to this last expression gives

$$B_{2,N} < 2a\bar{\theta}\ \exp\{-N(\delta')^2(2\sigma_{\theta}^{2}(h) + 2q\delta')^{-1}\}$$
$$\leqq 2a\ \exp\{-N(\delta')^2(2\bar{\sigma}^2(h) + 2q\delta')^{-1}\} \quad .$$

This exponential bound is independent of $\theta \in \Omega_\infty$ and hence $\lim_{N\to\infty} N^{1/2}B_{2,N} = 0$ uniformly in $\theta \in \Omega_\infty$.

This last result together with (28), when substituted into (25) implies $B_N = o(N^{-1/2})$ uniformly in $\theta \in \Omega_\infty$. A similar argument holds for C_N amd (ii) is proved. The theorem now follows by (i), (ii), and inequality (13).

If the estimate h is essentially bounded by M, then the conditions of Theorem 3 are met by taking $q = 3^{-1}M$. The estimate h in the example following Theorem 2 is an unbounded estimate satisfying the conditions of Theorem 3.

Theorem 4.

Let $h \in \mathcal{E}$ and $|h(u)| \leqq M$ a.e. μ. If (II) holds, then $R(\theta,t^*_h) - \phi(\bar{\theta}) = O(N^{-1})$ uniformly in $\theta \in \Omega_\infty$.

Proof. We bound the terms A_N, B_N and C_N in inequality (13). Expressing the term A_N in the integral form below and bounding in accord with (23) (which flows from assumption (II)), we obtain a uniform bound for A_N as follows:

$$A_N = P_\theta \int \{(z-\bar{\theta})([\bar{\theta} \leqq z < \bar{h}] - [\bar{h} \leqq z < \bar{\theta}])\} \frac{d\mu Z^{-1}}{d\lambda}(z)\, dz$$

$$\leqq (a+b)\, K' P_\theta \int (z-\bar{\theta})[\bar{\theta} \leqq z < \bar{h}]\, dz$$

$$= N^{-1}(a+b)\tfrac{1}{2}K'\sigma_\theta^2(h)$$

$$= N^{-1}(a+b)\tfrac{1}{2}K'\overline{\sigma^2}(h) \quad .$$

The term B_N can be treated in a similar manner after first bounding $\bar{h}^{(\alpha)} = \bar{h} + (h(u)-h(x_\alpha))N^{-1}$ from below by $\bar{h}-2MN^{-1}$ for each $\alpha \in I_1$ and then use assumption (II) to obtain,

$$B_N = a\bar{\theta}P_\theta P_1'[\bar{h}-N^{-1}(2M) \leqq Z < \bar{h}]$$

$$= a\bar{\theta}P_\theta \int [\bar{h}-N^{-1}(2M) \leqq z < \bar{h}] \frac{dP_1' Z^{-1}}{d\lambda}(z)\, dz$$

$$\leqq N^{-1}\, 2aK'M \quad .$$

In a similar manner, one has $C_N \leqq N^{-1}\, 2bK'M$.

Substituting these three upper bounds for A_N, B_N, and C_N respectively into inequality (13) yields an upper bound on the regret risk function for $t^*_{\bar{h}}$ given by $N^{-1}(a+b)\, K'(\overline{\sigma^2}(h) + 2M)$. Since this bound does not depend on $\theta \in \Omega_\infty$, the theorem is proved.

5. Examples Satisfying Theorem 3 or 4.

Two examples satisfying each of the Theorems 3 and 4 are given.

Example 1.

Let $U = (U_1,\ldots,U_n)$ be the generic random variable for the α^{th} problem. Assume $U_1,\ldots,U_n$ are independent identically distributed as either $G_o(t) = 1 - \exp\{-\omega_o t\}$, $\omega_o > 0$, $t \geqq 0$ or as $G_1(t) = 1 - \exp\{-\omega_1 t\}$, $\omega_1 > 0$, $t \geqq 0$. Furthermore, assume that $\omega_o < \omega_1 < 2\omega_o$. Let $g_o(t)$ and $g_1(t)$ be the Lebesque densities of $G_o(t)$ and $G_1(t)$. Then $Z(u)$ defined by (2) is given by

$$Z(u) = bf_o(u)$$

$$= \frac{b\,\Pi_{j=1}^{n} g_o(u_j)}{a\,\Pi_{j=1}^{n} g_1(u_j) + b\,\Pi_{j=1}^{n} g_o(u_j)}$$

$$= \{ab^{-1}(\omega_1\omega_o^{-1})^n \exp\{(\omega_o-\omega_1)\textstyle\sum_{j=1}^{n} u_j\} + 1\}^{-1} \quad .$$

The induced distributions $P_i Z^{-1}(z)$ for $i = 0,1$ are given by $P_i Z^{-1}(z) = \omega_i^n \int [Z(u)<z] \exp\{-\omega_i \sum_{j=1}^{n} u_j\} \Pi_{j=1}^{n} du_j$. Transforming this multiple integral by $v_k = \sum_{j=k}^{n} u_j$; $k = 1,\ldots,n$, which has Jacobian 1, followed by integration on the variable $v_n, v_{n-1}, \ldots, v_2$ yields for $i = 0,1$,

$$(29) \qquad P_i Z^{-1}(z) = \omega_i^n \Gamma^{-1}(n) \int_0^{(\omega_1-\omega_o)^{-1}\zeta(z)} v_1^{n-1} \exp\{-\omega_i v_1\}\, dv_1$$

where $\zeta(z) = \log\{(\omega_1\omega_o^{-1})^n ab^{-1} z(1-z)^{-1}\}$. For $i = 0$, transform this integral by means of the transformation $v_1 = (\omega_1-\omega_o)^{-1}\zeta(w)$ to obtain

$$P_o Z^{-1}(z) = C_o \int_C^z (1-w)^{(\omega_1-\omega_o)^{-1}(2\omega_o-\omega_1)}\, w^{(\omega_o-\omega_1)^{-1}\omega_1} [\zeta(w)]^{n-1} dw$$

where $C_o = \Gamma^{-1}(n)\{\omega_o(\omega_1-\omega_o)^{-1}(\omega_o\omega_1^{-1})^{\omega_o(\omega_1-\omega_o)^{-1}}\}^n (ba^{-1})^{\omega_o(\omega_1-\omega_o)^{-1}}$ and $C = b\omega_o^n(a\omega_1^n + b\omega_o^n)^{-1}$.

This integral expression immediately implies that $P_o Z^{-1}(z)$ is absolutely continuous with respect to Lebesgue measure λ, and we may define the following density

$$(30) \qquad \frac{dP_o Z^{-1}}{d\lambda}(z) = C_o(1-z)^{(\omega_1-\omega_o)^{-1}(2\omega_o-\omega_1)}\, z^{(\omega_o-\omega_1)^{-1}\omega_1} [\zeta(z)]^{n-1}$$

if $C \leqq z < 1$ and 0 otherwise.

Observe that the assumption $2\omega_o > \omega_1 > \omega_o$ implies that the factor $(1-z)^{(\omega_1-\omega_o)^{-1}(2\omega_o-\omega_1)}$ dominates the density (30) as $z \to 1$ and hence density (30) approaches 0 as $z \to 1$. This result implies that density (30) is continuous on the closed interval $[C,1]$, and hence the density

(30) is bounded on the closed interval [C,1] (and therefore on [0,1]). In a similar manner, it can be shown that

$$(31) \qquad \frac{dP_1Z^{-1}}{d\lambda}(z) = C_1(1-z)^{\omega_o(\omega_1-\omega_o)^{-1}} z^{(\omega_1-\omega_o)^{-1}(\omega_o-2\omega_1)} [\zeta(z)]^{n-1}$$

if $C \leq z < 1$ and 0 otherwise, where

$$C_1 = \Gamma^{-1}(n)\{\omega_1(\omega_1-\omega_o)^{-1}(\omega_o\omega_1{}^{-1})^{\omega_1(\omega_1-\omega_o)^{-1}}\}^n (ba^{-1})^{\omega_1(\omega_1-\omega_o)^{-1}} .$$

An argument similar to that following (30) shows that density (31) is bounded on [0.1]. Note that the assumption $2\omega_o > \omega_1$ is not necessary in showing (31) is bounded on [0,1].

Since (30) and (31) are bounded on [0,1], assumption (II) is verified and Theorem 4 holds for Example 1.

Example 2.

Same as Example 1 except assume that $\omega_1 \geq 2\omega_o$. Observe that the density (30) now approaches ∞ as $z \to 1$ and, hence, is unbounded on [0,1]. Therefore, the assumptions of Theorem 4 are violated. However, assumption (I) and, hence, Theorem 3 holds in this case by merely noting that (29) implies that $P_iZ^{-1}[Z=z] = 0$ for $i = 0,1$ if $C \leq z < 1$ (and therefore if $0 < z < 1$).

Example 3.

Let $U = (U_1,\ldots,U_n)$ be the generic random variable for the α^{th} problem. Assume $U_1,\ldots,U_n$ are independent identically distributed as either $G_o(t)$ or $G_1(t)$, where $G_i(t)$, for $i = 0,1$, is a normal distribution function with mean ω_i and standard deviation σ. Assume $\omega_1 < \omega_o$. Let $g_o(t)$ and $g_1(t)$ be the Lebesque densities of $G_o(t)$ and $G_1(t)$. Then $Z(u)$, defined by (2), is given by

$$Z(u) = bf_o(u)$$

$$= \frac{b \prod_{j=1}^{n} g_o(u_j)}{a \prod_{j=1}^{n} g_1(u_j) + b \prod_{j=1}^{n} g_o(u_j)}$$

$$= \{ab^{-1}c_1 \exp \{c_2 \sum_{j=1}^{n} u_j\} + 1\}^{-1} \quad ,$$

where $c_1 = \exp \{n(2\sigma^2)^{-1}(\omega_o^2-\omega_1^2).\}$ and $c_2 = (\omega_1-\omega_o)\sigma^{-2} < 0$.

Therefore, since $\sum_{j=1}^{n} U_j$ is the sum of n independent normals, the induced distributions $P_iZ^{-1}(z)$ for $i = 0,1$ are given by

$$P_iZ^{-1}(z) = P_i[\Sigma U_j < c_2^{-1} \log \{(ac_1z)^{-1} b(1-z)\}]$$

$$= \int_{-\infty}^{\zeta_i(z)} \Phi'(t)\, dt,$$

where $\zeta_i(z) = (n^{1/2}\sigma)^{-1}\{c_2^{-1} \log \{(ac_1z)^{-1} b(1-z)\} - n\omega_i\}$ and $\Phi'(t)$ is the density of $N(0,1)$.

For $i = 0,1$, transform the integrals by $t = \zeta_i(w)$ to obtain $P_iZ^{-1}(z) = \int_0^z \Phi'(\zeta_i(w))|\zeta_i'(w)|dw$. This integral expression immediately implies that $P_iZ^{-1}(z)$ is absolutely continuous with respect to Lebesque measure λ for $i = 0,1$. Since $|\zeta_i'(z)| = \{n^{1/2}\, \sigma|c_2|z(1-z)\}^{-1}$, the induced Lebesque densities are given by

$$\frac{dP_iZ^{-1}}{d\lambda}(z) = \{n^{1/2}\, \sigma|c_2|z(1-z)\}^{-1}\, \Phi'(\zeta_i(z)) \tag{32}$$

if $0 < z < 1$ and 0 otherwise for $i = 0,1$.

From the definition of $\zeta_i(z)$, we see that $\zeta_i(z) \to -\infty$ or ∞ according as $z \to 0$ or 1. Hence, $\Phi'(\zeta_i(z)) \to 0$ at an exponential rate as $z \to 0$ or 1 and thus $\Phi'(\zeta_i(z))$ is the dominant factor in (32) as $z \to 0$ or $z \to 1$. Therefore, (32) $\to 0$ as $z \to 0$ or $z \to 1$, for $i = 0,1$. Since the densities (32) are continuous on the open interval (0,1), the above argument shows that the densities (32) are continuous on the closed interval [0,1].

This in turn implies that these densities are bounded on $[0,1]$. Assumption (II) is thereby verified and Theorem 4 holds for Example 3.

Example 4.

Let $U = (U_1,\ldots,U_n)$ be the generic random variable for the α^{th} problem. Assume $U_1,\ldots,U_n$ are independent identically distributed random variables having distribution either $G_o(t)$ or $G_1(t)$. Furthermore, for $i = 0,1$ assume $G_i(t)$ is absolutely continuous with respect to Lebesque measure and has density $g_i(t) = c(\omega_i) \exp\{\omega_i T(t)\} h(t)$ where $T(t)$ is strictly monotone in t. Then $Z(u)$, defined in (2), is given by

$$Z(u) = bf_o(u)$$
$$= \left\{ab^{-1}\{c(\omega_1)c^{-1}(\omega_o)\}^n \exp\{(\omega_1-\omega_o)\textstyle\sum_{j=1}^n T(u_j)\} + 1\right\}^{-1} .$$

Note that the induced distributions $P_i Z^{-1}(z)$ for $i = 0,1$ are such that

(33) $$P_i Z^{-1}[Z=z] = P_i[\textstyle\sum_{j=1}^n T(U_j) = \zeta(z)]$$

for $0 < z < 1$, where

$$\zeta(z) = (\omega_1-\omega_o)^{-1} \log\{a^{-1}b(1-z)z^{-1}\}[c(\omega_o)\ c^{-1}(\omega_1)]^n .$$

With the aid of (33) we will show that $P_i Z^{-1}$ is continuous on $(0,1)$ for $i = 0,1$, and hence Theorem 3 holds.

Let $V(U_1,\ldots,U_n) = \sum_{j=1}^n T(U_j)$. The measurable transformation V from R^n into R induces a probability measure $P_i V^{-1}$, for $i = 0,1$, such that

(34) $$P_i[\textstyle\sum_{j=1}^n T(U_j) = \zeta(z)] = P_i V^{-1}[V = \zeta(z)]$$

for $0 < z < 1$.

Note that $P_i V^{-1}(v)$ is the distribution of the sum of n independent random variables $T_1,\ldots,T_m$, where $T_j = T(U_j)$, $j = 1,\ldots,n$. Each of n random variables T_j has, for $i = 0,1$, continuous induced distribution functions $P_i T_j^{-1}(t) = P_i[T(U_j) < t]$. Continuity follows since strict monotonicity of $T(\cdot)$ implies that $P_i T_j^{-1}[T_j = t] = P_i[T(U_j) = t]$ $= P_i[U_j = T^{-1}(t)] = 0$, for $i = 0,1$; $j = 1,\ldots,n$. Therefore, we conclude that $P_i V^{-1}(v)$, as the convolution of n continuous distribution functions, is continuous, for $i = 0,1$.

Hence, for $i = 0,1$, we obtain that $P_i V^{-1}[V = \zeta(z)] = 0$ for all z in (0,1). This, in turn, implies by (33) and (34) that $P_i Z^{-1}[Z=z] = 0$ for $i = 0,1$ and $z \in (0,1)$.

We have now exhibited a whole class of distributions for which assumption (I) and hence Theorem 3 are verified.

CHAPTER III

CONVERGENCE THEOREMS FOR THE GENERAL FINITE COMPOUND DECISION PROBLEM

1. Introduction.

In this chapter we shall extend Theorem 2 to the general finite compound decision problem of Chapter I, where the component problem has finite m x n loss matrix $(L(i,j))$. Counter-examples to the extensions of Theorems 3 and 4 are given. However, under a certain restriction on the loss matrix $(L(i,j))$, a theorem analogous to Theorem 4 is proved.

In Chapter I, we proposed the non-simple procedure $t'_{\bar{h}}$ defined by (1.12). To facilitate asymptotic study, we express the regret risk function of $t'_{\bar{h}}$ in the form (1) below. Let $p(\theta) = (p_0(\theta),\ldots,p_{m-1}(\theta))$ for $\theta \in \Omega_\infty$ be the empirical distribution of Ω. Recall that $t'_{p(\theta)}$ given by (1.8) with $\xi = p(\theta)$ is a simple decision procedure Bayes against $p(\theta)$. Hence, by (1.4) and (1.5) we may express $\phi(p(\theta)) = R(\theta,t'_{\bar{\theta}}) = (p(\theta),\rho(t'_{\theta}))$. Identify $t_\zeta = t'_{\bar{h}}$ in (1.15) of Lemma 5 and subtract the above term from the first term on the right-hand side of (1.15). Since Corollary 1 yields $(p(\theta),\rho(t'_{\bar{h}}) - \rho(t'_{p(\theta)}) \leqq (p(\theta)-\bar{h}, \rho(t'_{\bar{h}}) - \rho(t'_{p(\theta)})$, we then have

$$(1) \qquad R(\theta,t'_{\bar{h}}) - \phi(p(\theta) \leqq P_\theta(p(\theta)-\bar{h}, \rho(t'_{\bar{h}}) - \rho(t'_{p(\theta)})$$

$$+ N^{-1} \sum_{\alpha=1}^{N} \sum_{k\neq j} L^{kj}_{\theta_\alpha} P_\theta P_{\theta_\alpha} t'_{\bar{h}(\alpha),k}(U) t'_{\bar{h},j}(U).$$

When applying inequality (1) the first and second terms on the right-hand side of (1) will be denoted by A_N and B_N respectively.

2. Uniform Convergence Theorem of $O(N^{-1/2})$.

The following theorem generalizes Theorem 2 for an arbitrary $m \times n$ loss matrix.

Theorem 5.

If $h \in \mathcal{E}$ and $h_j \in L_3(P_i)$ for $i,j = 0,\ldots,m-1$, then $R(\theta,t'_{\bar{h}}) - \phi(p(\theta)) = O(N^{-1/2})$ uniformly in $\theta \in \Omega_\infty$.

Proof. In inequality (1) we show: (i) $A_N = O(N^{-1/2})$ uniformly in $\theta \in \Omega_\infty$ and (ii) $B_N = O(N^{-1/2})$ uniformly in $\theta \in \Omega_\infty$.

(i) By the Schwarz m-space inequality, we have,

$$(2) \qquad \begin{aligned} N^{1/2} A_N &\leqq N^{1/2} P_\theta |(p-\bar{h}, \rho(t'_{\bar{h}}) - \rho(t'_p)| \\ &\leqq N^{1/2} P_\theta \|\bar{h}-p\| \, \|\rho(t'_{\bar{h}}) - \rho(t'_p)\| \, . \end{aligned}$$

Let $\underline{L}_i = \min_j L(i,j)$ and $\overline{L}_i = \max_j L(i,j)$ and note that $\underline{L}_i \leqq \text{Range } \rho_i \leqq \overline{L}_i$. Then

$$(3) \qquad \begin{aligned} \|\rho(t'_{\bar{h}}) - \rho(t'_p)\|^2 &= \sum_{i=o}^{m-1} \{\rho_i(t'_{\bar{h}}) - \rho_i(t'_p)\}^2 \\ &\leqq \sum_{i=o}^{m-1} (\overline{L}_i - \underline{L}_i)^2 \\ &= \|\overline{L} - \underline{L}\|^2, \text{ where} \end{aligned}$$

$\overline{L} = (\overline{L}_o,\ldots,\overline{L}_{m-1})$ and $\underline{L} = (\underline{L}_o,\ldots,\underline{L}_{m-1})$.

Also, note that by the Schwarz integral inequality and Lemma 4,

$$(4) \qquad N^{1/2} P_\theta \|\bar{h}-p(\theta)\| \leqq \{N P_\theta \|\bar{h}-p(\theta)\|^2\}^{1/2} \leqq C.$$

Inequalities (3) and (4), when substituted into (2), imply $N^{1/2} A_N \leqq C\|\overline{L}-\underline{L}\|$. Hence, (i) is proved.

(ii) Let $I_i = \{\alpha | \theta_\alpha = i\}$, $i = 0,\ldots,m-1$. Let r_i be the rank of the covariance matrix of $h = (h_o,\ldots,h_{m-1})$ under the distribution P_i, $i=0,\ldots,m-1$. Fix i, j, k, $k < j$ and $\alpha \varepsilon I_i$, and let $d = L^{kj} f(u)$ and $\ell_{kj} = t'_{\bar{h}(\alpha),k}(u)\, t'_{\bar{h},j}(u)$. Let $h - \varepsilon_i = TZ$ with Z_ℓ an orthonormal basis for the subspace of $L_2(P_i)$ generated by the functions $h_\ell - \delta_{\ell i}$, where $\varepsilon_i = (\delta_{oi},\ldots,\delta_{m-1,i})$ and $\delta_{\ell i}$ is the Kronecker δ. Abbreviate $\|T'd\|^{-1} T'd$ to g, where T' is the transpose of the matrix T.

Observe that Z_ℓ, as a linear combination of the functions h_j in $L_3(P_i)$, is in $L_3(P_i)$, $\ell = 1,\ldots,r_i$. Also, since $\{Z_\ell | \ell = 1,\ldots,r_i\}$ is a set of orthonormal functions in $L_2(P_i)$, we have,

(5) $$P_i(Z,g)^2 = ||g||^2 = 1$$

and,

(6) $$P_i\|Z\|^2 = r_i.$$

We note that since $\ell_{kj} \leqq [0 < (\bar{h},d) \leqq (\bar{h}-\bar{h}^{(\alpha)},d)]$, $\ell_{kj} + \ell_{jk} \neq 0$ implies $r_i > 0$ and $T'd \neq 0$. Suppose $\ell_{kj} + \ell_{jk} \neq 0$ and $Np_i > 1$. Then, conditionally on u, x_α and all x_ω, $\omega \notin I_i$, the sum $\sum_{\omega\neq\alpha,\omega\varepsilon I_i}(Z(x_\omega),g)$ falls into an interval of length $|(Z(x_\alpha)-Z(u),g)|$. Hence, a B-E approximation to this conditional probability of $\ell_{kj} + \ell_{jk}$ yields a bound, $(Np_i-1)^{-1/2}\{\Phi'(0)|(Z(x_\alpha)-Z(u),g)| + 2bP_i|(Z,g)|^3\}$, after simplification by (5). Taking the bound on this conditional probability to be 0 if $\ell_{kj} + \ell_{jk} = 0$ and 1 if $Np_i = 1$ and weakening the $P_\theta \times P_i$ integral in this bound by the Schwarz r_i-space and integral inequalities, the triangle inequality, and (6) used to obtain $P_\theta P_i|(Z(X_\alpha) - Z(U),g)| \leqq P_\theta P_i\|Z(X_\alpha) - Z(U)\| \leqq 2\{P_i\|Z\|^2\}^{1/2} = 2\, r_i^{1/2}$, we have if $Np_i \geqq 1$,

(7) $$P_\theta P_i(\ell_{kj} + \ell_{jk}) \leqq \min\{1, (Np_i - 1)^{-1/2} C_i\},$$

where $C_i = \Phi'(0)\, 2(r_i)^{1/2} + 2\beta P_i \|Z\|^3$. If $r_i = 0$, (7) holds with $C_i = 0$ and $0 \cdot \infty = 0$.

Observe that inequality (2.14) implies that

$$N^{1/2} p_i \min\{1, |Np_i - 1|^{-1/2} C_i\} \leqq p_i^{1/2}(1 + C_i^2)^{1/2} \text{ for all } i.$$

Hence, since $\sum_{i=0}^{m-1} p_i = 1$, we have by the Schwarz m-space inequality

$$(8) \qquad \sum_{i=0}^{m-1} N^{1/2} p_i \min\{1, |Np_i - 1|^{-1/2} C_i\} \leqq (m + \|C\|^2)^{1/2} \quad .$$

Noting that $B_N \leqq N^{-1} \sum_{\alpha=1}^{N} \sum_{k<j} |L_{\theta_\alpha}^{kj}| P_\theta P_{\theta_\alpha} (\ell_{kj} + \ell_{jk})$, we see that (7) and (8) imply

$$(9) \qquad N^{1/2} B_N \leqq \binom{n}{2} L (m + \|C\|^2)^{1/2} \quad ,$$

where $L = \max_{i,j,k} |L_i^{kj}|$.

Equation (9) implies (ii), which together with (i) and inequality (1) completes the proof.

3. Sufficient Conditions for a Theorem of Higher Order.

In this section we shall examine certain sufficient conditions which allow a generalized analogue of Theorem 4 in Chapter II. Two types of sufficient conditions are imposed: a certain continuity assumption relating to the class of probability measures $\{P_0, \ldots, P_{m-1}\}$, and a condition on the m x n component loss matrix $(L(i,j))$. The continuity assumption is a "natural" extension of the sufficient condition (II) of Theorem 4 in Chapter II. That an additional condition is needed on the loss matrix will be illustrated by two examples.

Consider the following example, which illustrates that, regardless of what continuity assumptions are imposed on a class $\{P_0, \ldots, P_{m-1}\}$

satisfying a mild regularity assumption (see (9) below), a uniform convergence theorem of rate faster than $O(N^{-1/2})$ is unobtainable for a certain loss matrix.

<u>Example.</u> Let $n = 2$ and $h = (h_o,\ldots,h_{m-1}) \in \mathcal{E}$ such that $h_j \in L_3(P_i)$ for $i,j = 0,\ldots,m-1$. Let $I = (I_+,I_o,I_-)$ be a proper partition of $\{0,\ldots,m-1\}$ according to $L_i^{10} >, =$ or < 0. Define $w_u(v) = (L^{10}h(v),f(u))$. Note that $w_u \in L_3(P_i)$ for $i = 0,\ldots,m-1$. Assume there exists $i \in I_o$, $i' \in I_+ \bigcup I_-$ such that,

$$(9) \qquad P_{i'}\,[\sigma_i^{\,2}(w_U) > 0] > 0.$$

Without loss of generality, we may assume $i' \in I_+$. Existence of a class $\{P_o,\ldots,P_{m-1}\}$ satisfying (9) can be assured by taking common support $S = \{u|f_i(u) > 0\}$ for all i, and noting that under this assumption condition (9) is equivalent to $L^1 \neq L^0$.

Consider now $\theta \in \Omega_\infty$ such that $0 < \gamma \leq N^{1/2}\, p_{i'} \leq \delta \leq \infty$ and $p_i = 1 - p_{i'}$ for all N sufficiently large. Fix α such that $\theta_\alpha = i'$ and define the set $E = \{\sum_{\ell=1}^{N} w_{x_\alpha}(x_\ell) < 0\}$. Define $s_N^2(u) = Np_i\, \sigma_i^{\,2}(w_u) + (Np_{i'}-1)\, \sigma_{i'}^{\,2}(w_u)$ and $K_N(u) = -s_N^{-1}(u)\,\{w_u(u) + (Np_{i'}-1)\, L_{i'}^{10}\, f_{i'}(u)\}$. Then, by a B-E approximation applied conditionally on $X_\alpha = u$, we have

$$(10) \qquad P_\theta[E|X_\alpha = u] \geqq Y_N^+(u) \quad ,$$

where $Y_N(u) = \Phi(K_N(u)) - \beta s_N^{-3}(u) \sum_{\ell\neq\alpha} P_{\theta_\ell} |w_u - P_{\theta_\ell} w_u|^3$.

Note that on $\{u|\sigma_i^{\,2}(w_u) > 0\}$, $N^{-1}s_N^2(u) \sim \sigma_i^{\,2}(w_u) > 0$, and hence on this set $\underline{\lim}\, K_N(u) \geq C(u)$, where $C(u) = -\delta\, L_{i'}^{10}\, f_{i'}(u)\, \sigma_i^{-1}(w_u)$. Thus, since $\underline{\lim}\, Y_N^+ \geqq (\underline{\lim}\, Y_N)^+$ and $\Phi(\cdot)$ is an increasing function, we have

$$(11) \qquad \underline{\lim}\, Y_N^+(u) \geqq \Phi(C(u)) \text{ on } \{\sigma_i^{\,2}(w_u) > 0\} \quad .$$

Therefore, Fatou's Lemma, (10), and (11) imply,

(12) $$\underline{\lim}\, P_\theta[E] = \underline{\lim}\, P_{i'} P_\theta[E|X_\alpha = u]$$
$$\geqq P_{i'}\, \underline{\lim}\, P_\theta\, [E|X_\alpha = u] \geqq C,$$

where $C = P_{i'}\, [\sigma_i^2(w_U) > 0]\, \Phi(C(U)) > 0$.

Finally, since L^{10} is optimal against both i and i', we see that

(13) $$\underline{\lim}\quad N^{1/2}\{R(\theta, t_{\overline{h}}') - \phi(p(\theta))\}$$
$$\geqq \underline{\lim}\, N^{-1/2} \sum_{\alpha=1}^{N} P_\theta L_{\theta_\alpha}^{10}\, t_{\overline{h},1}'(X_\alpha)$$
$$= \underline{\lim}\, N^{1/2}\; p_{i'}\; L_{i'}^{10}\; P_\theta[E]$$
$$\geqq \gamma\, L_{i'}^{10}\, C > 0.$$

Inequality (13) contradicts the possibility of a uniform convergence theorem of order greater than $O(N^{-1/2})$ in the general finite compound decision problem with arbitrary loss matrix.

Consider now the following condition (C) on the loss matrix $(L(i,j))$. Let $I_{kj} = \{i | L_i^{kj} = 0\}$. The condition is:

(C) For all j,k $(j \neq k)$ and $i \in I_{kj}$, there exists an $\ell = \ell(i,j,k)$ such that $L_i^{j\ell} > 0$ and $L_{i'}^{j\ell} \geqq 0$ on I_{kj}.

Note that condition (C) is violated in the example above for all $i \in I_o$. With this added restriction (C) we will obtain a uniform convergence theorem for the regret risk function of $O(N^{-1})$. The sufficiency of (C), together with the continuity assumption (II') (or II") below, will be seen in the proof of Theorem 6. A certain degree of necessity for this condition is shown by the above example and is demonstrated more clearly by the example in section 3.5.

We mention here three important cases in which (C) is satisfied. All three cases are concerned with the discrimination problem in which

$m=n$ and $L(i,j)=0$ or >0 according as $i=j$ or $i\neq j$. The three cases are:

(i) Let $m = 2$ or 3. This case reduces to the problem of Chapter II for $m = 2$.

(ii) Define $L(i,j) = a(1-\delta_{ij})$, where δ_{ij} is the Kronecker δ. Condition (C) is satisfied by choosing $\ell(i,j,k) = i$.

(iii) Let $w(t)$ be a strictly increasing function on $[0,\infty)$ with $w(o) = 0$. Define $L(i,j) = w(|i-j|)$. Since $L_i^{kj} = 0$ for $j \neq k$ implies $i > j$ and $i < k$ or $i < j$ and $i > k$, condition (C) is satisfied by choosing $\ell(i,j,k) = i$.

We now examine the sufficient condition to be imposed on the class $\{P_o,\ldots,P_{m-1}\}$. Let μ be some dominating measure for the P_i's and define $f = (f_o,\ldots,f_{m-1})$, where f_i is the density of P_i with respect to μ. Let $P_i f^{-1}$ denote the probability measure induced under the measurable transformation $u \to f(u)$. Note that $P_i f^{-1}$ is a probability measure on $(R^m, \mathcal{B}^m)$, where $\mathcal{B}^m$ is the σ-field of Borel sets on Euclidean m-space. Let λ_m denote m-dimensional Lebesque measure. Define B_j in $\mathcal{B}^m$, $j = 0,\ldots,m-1$ as

(14) $$B_j = B_j(v,a,b) \leq \{0 \leq (v,f) \leq a,\ 0\leq f_j \leq b,\ 0\leq f_i \leq K,\ i\neq j\},$$

where $|v| = 1$, $a \geq 0$, $b \geq 0$. Consider now the following condition on $\{P_o,\ldots,P_{m-1}\}$:

(II') There exists a measure μ dominating the P_i's and finite constants K, K' such that $P_i f^{-1}[B_j] \leq K'\lambda_m[B_j]$ for all i,j,v,a, and b with $v_j(b-K) = 0$ and B_j of the form (14).

This condition is by no means an obvious generalization of condition (II) of Theorem 3. However, let $P. = \sum_{i=o}^{m-1} P_i$ and let $Z_i(u)$ be the density of P_i with respect to $P.$. Define $\overline{Z}$ as the measurable transformation $u \to (Z_1(u),\ldots,Z_{m-1}(u))$ and $P_i\overline{Z}^{-1}$ the induced measure on $(R^{m-1}, \mathcal{B}^{m-1})$ under $\overline{Z}$. Let λ_{m-1} denote m-1 dimensional Lebesque measure. Then we can state the following "natural" extension of condition (II) as:

(II") For $i = 0,\ldots,m-1$, $P_i\overline{Z}^{-1}$ is absolutely continuous with respect to λ_{m-1} and for some $K'' < \infty$,

$$\frac{d\, P_i\overline{Z}^{-1}}{d\lambda_{m-1}} \leqq K'' \quad . \tag{15}$$

Condition (II') is seen to be equivalent to condition (II) of Chapter II by observing for $m = 2$, $P_i[\overline{Z}(U) < z] = P_i[Z(U) > C(z)]$, where $C(z) = \{b + (a-b)z\}^{-1}b(1-z)$ and $Z(u)$ is defined by (2.2).

It can now be seen that condition (II') generalizes condition (II) in the sense that condition (II"), which is equivalent to (II) for $m = 2$, implies (II') when $\mu = P.$ in (II'). For the proof of this statement, see Appendix 1.

We now give an example which fulfills condition (II').

<u>Example.</u> Let $U = (U_o,\ldots,U_{m-1})$ be the generic random variable for the component problem. Define, for $i = 0,\ldots,m-1$, the probability measures P_i having densities with respect to λ_m given by $f_i(u) = 2\, u_i$ if $u \in [0,1]^m$. If we let $P_i f^{-1}(f_o,\ldots,f_{m-1})$ be the distribution function corresponding to the induced probability $P_i f^{-1}$, then $P_i f^{-1}(f_o,\ldots,f_{m-1}) = 2^{-(m+1)}\, (\Pi_{j=o}^{m-1} f_j)\, f_i$ on $f \in [0,2]^m$. Hence, $P_i f^{-1}$ is absolutely continuous with respect to λ^m and has λ_m-density $2^{-m}\, f_i$ on $[0,2]$, which is bounded by $2^{-\,m+1}$ on $[0,2]^m$. Therefore, with $K' = 2^{-\,m+1}$, $P_i f^{-1}[B] \leqq K'\, \lambda_m[B]$ for all Borel sets B on R^m;

and, hence, in particular for the sets B_j of condition (II') with $K = 2$.

This example may be generalized. Let P_i be a probability measure on m-space with λ_m-density $g_i(u)$. Choose, if possible, a measure μ such that $P_i << \mu << \lambda_m$ with $h(u)$ as the λ_m-density of μ and such that $u \to f(u) = g(u)/h(u)$ is a 1-1 map from $\{u|f(u) \neq 0\}$ into $[0,K]^m$ having Jacobian $J(u(f)/f)$ with $h(u(f))\, J(u(f)/f)$ bounded by K_0. Then, on the range of f, we have

$$(16) \qquad \frac{dP_i f^{-1}}{d\,\lambda_m}(f_o,\ldots,f_{m-1}) = f_i(u(f))h(u(f))J(u(f)/f)$$
$$\leqq K\, K_0 < \infty \quad .$$

But (16) implies that $P_i f^{-1}[B] \leqq K\, K_0\, \lambda_m[B]$ for all Borel sets B on R^m; and hence, in particular for the sets B_j of condition (II').

In the example given above, $\mu = \lambda_m$, $h(u) = 1$, and $g_i(u) = f_i(u) = 2\, u_i$ for $u \in [0,1]^m$ with $K = 2$ and $K_0 = 2^{-m}$. Another example in which μ plays a more dominant role is with $h(u) = 2^m \;\Pi_{j=o}^{m-1} u_j$, $g_i(u) = 2^{m-1}\, 3\, u_i \;\Pi_{j=o}^{m-1} u_j$, and $f_i(u) = 2^{-1}\, 3\, u_i$ for $u \in [0,1]^m$, and with $K_0 = 4^m\, 3^{-m}$ and $K = 2^{-1}\, 3$.

4. Uniform Convergence Theorem of $O(N^{-1})$.

Before stating and proving Theorem 6, we shall prove the following useful lemma.

Lemma 6.

For sets $B_j = B_j(v,a,b)$ of the form (14), $\lambda_m[B_j] \leqq a\, b\, K^{m-2}$ if $|v_j| < 1$.

Proof. Let $\ell \neq j$ be such that $|v_\ell| = 1$. The lemma follows from the transformation $y_\ell = (v,f)$ and $y_k = f_k$, $k \neq \ell$, which has unit Jacobian.

Theorem 6.

If (C) and (II') hold and $h \in \mathcal{E}$ such that $|h_i(u)| \leqq M$ a.e. μ for $i = 0,\ldots,m-1$, then $R(\theta,t_{\bar{h}}') - \phi(p(\theta)) = O(N^{-1})$ uniformly in $\theta \in \Omega_\infty$.

Proof. We show in inequality (1) that: (i) $A_N = O(N^{-1})$ uniformly in $\theta \in \Omega_\infty$ and (ii) $B_N = O(N^{-1})$ uniformly in $\theta \in \Omega_\infty$.

(i) By noting $((p_i - \bar{h}_i)L_i,\ t_{\bar{h}}'(u) - t_p'(u))$ is the difference of two simple functions, we see that the first term on the right-hand side of (1) can be written as

$$(17) \qquad A_N = P_\theta(p-\bar{h},\ \rho(t_{\bar{h}}') - \rho(t_p')) = \textstyle\sum_{j\neq k} D_N(k,j),$$

where $D_N(k,j) = P_\theta \sum_{i=o}^{m-1} (p_i-\bar{h}_i)L_i^{kj}\ P_i t_{\bar{h},k}'(U)\ t_{p,j}'(U)$. Without loss of generality, we may assume $p_i > 0$ for all $i = 0,\ldots,m-1$ in A_N, since, if $p_i = 0$, the term $p_i(\rho_i(t_{\bar{h}}') - \rho_i(t_p')) = 0$ could be eliminated prior to use of Corollary 1 in (1).

Fix i, j, k, and observe that for $\ell = 0,\ldots,m-1$,

$$(18) \qquad t_{\bar{h},k}'(u)\ t_{p,j}'(u) \leqq [0 \leqq (pL^{kj},f(u)) \leqq m^{\frac{1}{2}}LK\|p-\bar{h}\|][\textstyle\sum_{i\in I_{kj}} p_i L_i^{j\ell} f_i(u) \leqq \sum_{i\notin I_{kj}} p_i L_i^{\ell j} f_i(u)].$$

Consider the following two cases.

Case 1. Let $\max_{i\notin I_{kj}} p_i \geqq m^{-1}$. Bound the second factor on the right-hand side of (18) by unity and note that condition (II') and Lemma 6 applied to the remaining factor with $v = |pL^{kj}|^{-1}\ pL^{kj}$, $a = |pL^{kj}|^{-1}\ m^{1/2}\ LK\ \|p-\bar{h}\|$ and $b = K$ yields

(19) $$P_i t'_{\bar{h},k}(U)\ t'_{p,j}(U) \leq a\ b\ K^{m-2}\ K' \leq K_m \,\|p-\bar{h}\|,$$

where $K_m = m^{3/2}\ L\ L_0^{-1}\ K^m\ K'$ with $L_0 = \min_{i,j,k}\ \{|L_i^{kj}|\ |\ L_i^{kj} \neq 0\}$.

Case 2. Let $0 < \max_{i \notin I_{kj}} p_i < m^{-1}$. Then there exists an $\omega \in I_{kj}$ such that $p_\omega \geq m^{-1}$. Therefore, by condition (C), $L_{i'}^{j\ell} \geq 0$ on I_{kj} and $L_\omega^{j\ell} > 0$ for some ℓ. For such an ℓ, we have $\sum_{i \in I_{kj}} p_i L_i^{j\ell} f_i(u) \geq p_\omega L_\omega^{j\ell} f_\omega(u) \geq 0$ and $|L_i^{\ell j}| \leq |L_i^{kj}|\ L\ L_0^{-1}$ for $i \notin I_{kj}$. Hence, the second factor on the right-hand side of inequality (18) is bounded by $[0 \leq p_\omega L_\omega^{j\ell} f_\omega(u) \leq (m-1)L\ L_0^{-1} K\,|pL^{kj}|]$. With this bound in (18), condition (II') and Lemma 6 applied to $B_\omega(v,a,b)$ with $v = |pL^{kj}|^{-1}\ pL^{kj}$, $a = m^{\frac{1}{2}} LK\ \|p-\bar{h}\|\ |pL^{kj}|^{-1}$, and $b = (p_\omega\ L_\omega^{j\ell}\ )^{-1}(m-1)\ L\ L_0^{-1}\ K\ |pL^{kj}|$, we have

(20) $$P_i\ t'_{\bar{h},k}(U)\ t'_{p,j}(U) \leq ab\ K^{m-2}\ K' \leq K'_m\ \|p-\bar{h}\|,$$

where $K'_m = m^{3/2}(m-1)(L\ L_0^{-1})^2\ K^m\ K'$ is obtained by noting that $p_\omega\ L_\omega^{j\ell}\ \geq m^{-1}\ L_0$.

Observing that $K'_m \geq K_m$ for $m \geq 2$, substitute the bound in (20) into the term $D_N(k,j)$ for both case 1 and case 2 to obtain with the aid of the Schwarz m-space inequality $D_N(k,j) \leq m^{1/2}\ L\ K'_m\ P_\theta\|\bar{h}-p\|^2$. Hence, Lemma 4 and equality (17) imply $A_N \leq \{n(n-1)m^{1/2}LK'_m C^2\}N^{-1}$, from whence (i) follows.

(ii) Fix i,j,k and define $E = \{0 \leq (\bar{h}L^{kj}, f(u)) \leq \alpha_1 N^{-1}\}$, $F_\ell = \{0 \leq (\bar{h}L^{\ell j}, f(u))\}$, $\ell = 0,\ldots,m-1$, and $\alpha_1 = 2mMLK$. Note that by the definition of $t'_{\bar{h}(\alpha),k}(u)$ and $t'_{\bar{h},j}(u)$ we have $t'_{\bar{h}(\alpha),k}(u)\ t'_{\bar{h},j}(u) \leq [E]\ [F_\ell]$ for $\alpha = 1,\ldots,N$. Hence,

$$(21) \qquad N^{-1} \sum_{\alpha \varepsilon I_i} L^{kj}_{\theta_\alpha} P_\theta P_{\theta_\alpha} t'_{\bar{h}(\alpha),k}(u)\, t'_{p,j}(u)$$

$$\leqq |p_i L_i^{kj}|\, P_\theta P_i\, [E]\,[F_\ell].$$

We now consider bounding the right-hand side of (21) in two cases.

Case 1. Let $\max_{i \notin I_{kj}} p_i \geqq m^{-1}$. Define the set $A=\{\|\bar{h}-p\| \leqq (2m)^{-1}\}$. Note that on A, $|\bar{h}L^{kj}| \geqq L_0(2m)^{-1}$, and hence by condition (II') and Lemma 6 we have, $P_\theta P_i[E][A] = P_\theta[A]P_i[E] \leqq (2m\, L_o^{-1}\, K^{m-1}\, K'\alpha_1)\, N^{-1}$. Also, we have by Tchebichev's inequality and Lemma 4, $P_\theta(1 - [A]) \leqq 4\, m^2\, P_\theta\, \|\bar{h}-p\|^2 \leqq 4(mC)^2\, N^{-1}$. Hence, with $\alpha_2 = 2m\, L_0^{-1} K^{m-1} K'\alpha_1 + 4(mC)^2$, it follows that

$$(22) \qquad |p_i L_i^{kj}|\, P_\theta P_i([E][F_\ell]) \leqq p_i L \alpha_2\, N^{-1}.$$

Case 2. Let $\max_{i \notin I_{kj}} p_i < m^{-1}$. Then there exists an $\omega \varepsilon I_{kj}$ such that $p_\omega \geqq m^{-1}$. By condition (C), there exists an $\ell = \ell(\omega,j,k)$ such that $L_\omega^{j\ell} > 0$ and $L_{i'}^{j\ell} \geqq 0$ on I_{kj}.

Observe that $|p_i L_i^{kj}| \leqq |p_i - \bar{h}_i|\, L + |\bar{h}\, L^{kj}|$. Then, since (II') and Lemma 6 imply, for $|\bar{h}\, L^{kj}| > 0$, $P_i[E] \leqq \alpha_3\, |\bar{h}\, L^{kj}|^{-1}\, N^{-1}$ where $\alpha_3 = K^{m-1}\, K'\alpha_1$, we have

$$(23) \qquad |p_i L_i^{kj}|\, P_i([E][F_\ell]) \leqq L|p_i - \bar{h}_i|\, P_i([E][F_\ell]) + \alpha_3\, N^{-1}.$$

With $\ell=\ell(\omega,j,k)$ and observing that $\sum_{i \notin I_{kj}} \bar{h}_i L_i^{\ell j} f_i(u) \leqq mLL_0^{-1}K|\bar{h}L^{kj}|$ and that $\sum_{i \varepsilon I_{kj}} (\bar{h}_i - p_i) L_i^{\ell j} f_i(u) \leqq m^{1/2} LK\, \|p-\bar{h}\|$ on F_ℓ, we obtain the set inclusion, $F_\ell \subset \{0 \leqq p_\omega L_\omega^{j\ell} f_\omega(u) \leqq (mL_0^{-1}|\bar{h}L^{kj}| + m^{1/2}\|p-\bar{h}\|)LK\}$. Let $G = \{|\bar{h}L^{kj}| < N^{-1/2}\}$. Then, since $|p_\omega L_\omega^{j\ell}| \geqq m^{-1}\, L_0$ we have on G, (II') and Lemma 6 implying by the above set inclusion $P_i[F_\ell] \leqq (mL_0^{-1}\, N^{-1/2} + m^{1/2}\, \|p-\bar{h}\|)\, mLL_0^{-1}K^m K'$, while on the complement of G,

$P_i[E] \leq K^{m-1} K'\alpha_1 N^{-1/2}$. Hence, we have in the term $P_i([E][F_\ell])$ for $\ell = \ell(\omega,j,k)$,

$$(24) \qquad P_i([E][F_\ell]) \leq \alpha_4 N^{-1/2} + \alpha_5 \|p-\bar{h}\|,$$

where $\alpha_4 = K^{m-1} K' \{(m L_0^{-1})^2 KL + \alpha_1\}$ and $\alpha_5 = m^{3/2} L L_0^{-1} K^m K'$.

To complete the proof for the term B_N, substitute (24) into the first term on the right-hand side of (23), sum the P_θ integral of this bound plus the bound in (22) over all $i,j,k,(j\neq k)$, and use Schwarz' inequality to bound $\sum_{i=o}^{m-1} |p_i-\bar{h}_i| \leq m^{1/2} \|p-\bar{h}\|$. The resulting inequality from the definition of B_N and inequality (21) is

$$(25) \qquad B_N \leq n(n-1) \{(L\alpha_2 + m\alpha_3) N^{-1} + m^{1/2}L(\alpha_4 N^{-1/2} P_\theta\|\bar{h}-p\| + \alpha_5 P_\theta\|\bar{h}-p\|^2)\} .$$

From (25) we see that $B_N = O(N^{-1})$ uniformly in $\theta \in \Omega_\infty$, since by Schwarz' inequality and Lemma 4, $N^{-1/2} P_\theta \|\bar{h}-p\| \leq N^{-1}C$ and $P_\theta \|\bar{h}-p\|^2 \leq C^2N^{-1}$. Therefore, (ii) is proved, which together with (i) and inequality (1) completes the proof.

5. Counter-example to Theorem 6 when (C) is Violated.

The example given in this section shows that even in the discrimination case ($L(i,j) >$ or $= 0$ as $i \neq j$ or $i = j$, and $m = n$) and with condition (II') satisfied, a violation of condition (C) prohibits a uniform convergence theorem of order greater than $O(N^{-1/2})$. This example together with the first example in section 3.3 exhibits that condition (C), although maybe not a necessary condition for Theorem 6, is at least not an unwarranted assumption on the loss matrix $(L(i,j))$.

<u>Example.</u> Let $m = n = 4$ and assume $\{P_0,\ldots,P_3\}$ satisfy condition (II'). Let $h \varepsilon \mathcal{E}$ with $|h_i(u)| \leqq M$ a.e. μ for $i = 0,\ldots,3$. Suppose the component loss matrix is given by:

$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 2 & 1 \\ 1 & 7 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} .$$

Note that in this example all conditions of Theorem 6 are met except condition (C) which is violated when $(k,j) = (0,3)$ or $(3,0)$. As will be seen, the conclusion of Theorem 6 is not true for this example.

To facilitate construction of the example, choose distributions with densities $f_i(u)$ having common support set $S = \{u | f_i(u) > 0\}$ for $i = 0,1,2,3$, and such that $\frac{1}{2} f_2(u) \leqq f_1(u) \leqq 2 f_2(u)$ on S. Furthermore, assume that $\kappa \leqq f_1(u) \leqq K$ on S for some constants κ, K, $0 < \kappa < K < \infty$.

To see that this class of examples is non-empty, let $f_i(u) = 3^{-1}2u_i$ on $[1,2]^4$ with $\mu = \lambda_4$, $\kappa=3^{-1}2, K=3^{-1}4$. Then, $S = [1,2]^4$ and $u_2 \leqq 2u_1 \leqq 4u_2$ on S. That condition (II') is satisfied follows by analogy with the example satisfying (II') given in section 3.3 with m=4.

Now choose $\theta' \varepsilon \Omega_\infty$ such that for N sufficiently large, $0 < \gamma \leqq N^{1/2}\, p_o(\theta') \leqq \delta < \infty$, $p_3(\theta') = 0$, and $2p_2(\theta') + \varepsilon \leqq p_1(\theta') \leqq 3p_2(\theta') - \varepsilon$, $0 < \varepsilon < \frac{1}{2}$. By the choice of θ', $t'_{p(\theta'),0}(u) = 1$ a.e.μ. Hence, $R(\theta',t_{\bar{h}}') - \phi(p(\theta')) = N^{-1} \sum_{\alpha=1}^{N} P_{\theta'} \sum_{k=1}^{3} L_{\theta_\alpha}^{k0}\, t_{\bar{h},k}(X_\alpha)$. Note that condition (C) is satisfied for $k = 1,2$ and hence by the proof of Theorem 6 and the fact that $p_3(\theta') = 0$, we see

(26) $$R(\theta',t'_{\bar{h}}) - \phi(p(\theta')) = N^{-1} \sum_{\alpha\varepsilon I_0} P_{\theta'} t'_{\bar{h},3}(X_\alpha) + O(N^{-1}).$$

Fix $\alpha \varepsilon I_0$ and define $E = [\bar{h}_0 f_0(X_\alpha) - \bar{h}_3 f_3(X_\alpha) < 0]$. Conditionally on $X_\alpha = u$, apply a B-E approximation to the sum $\sum_{\ell\neq\alpha}(h_0(X_\ell)f_0(u) - h_3(X_\ell)f_3(u))$ and let $\sigma(u) = \min_{i=0,1,2} \sigma_i(h_0(V)f_0(u) - h_3(V)f_3(u)) > 0$ on S to obtain,

(27) $$\underline{\lim}\ P_{\theta'}[E \mid X_\alpha = u] \geq \Phi(-\delta f_0(u)\ \sigma^{-1}(u)) \text{ on } S.$$

Now, observe the following pointwise inequalities:

(28) $$0 \leq [E] - t_{\bar{h},3}(X_\alpha) \leq 1 - [(\bar{h},L^{31}f(X_\alpha)) < 0]\ [\bar{h},L^{32}f(X_\alpha) < 0]$$
$$\leq [(\bar{h},L^{31}f(X_\alpha)) \geq 0] + [(\bar{h},L^{32}f(X_\alpha)) \geq 0].$$

Then, with $(\bar{h},L^{3k}f(X_\alpha)) \leq \|\bar{h}-p(\theta')\|\ 12\ K + (p(\theta'),L^{3k}f(X_\alpha))$ for $k=1,2$, while our choice of the f_i's and θ' implies $(p(\theta'),L^{k3}f(X_\alpha)) \geq \varepsilon f_1 \geq \varepsilon\kappa$, we see that (28) together with Tchebichev's inequality and Lemma 4 imply

(29) $$0 \leq P_{\theta'}\{[E] - t_{\bar{h},3}(X_\alpha)\} \leq 2\ P_{\theta'}[\|\bar{h}-p(\theta')\|\ 12\ K \geq \varepsilon\ \kappa] \leq \alpha_0\ N^{-1},$$

where $\alpha_0 = 288\ (KC)^2\ (\varepsilon\kappa)^{-2}$.

Thus, we have $P_\theta t'_{\bar{h},3}(X_\alpha) = P_\theta[E] + O(N^{-1})$ from whence it follows by (26) and (27) that

(29) $$\underline{\lim}\ N^{1/2}\ (R(\theta',t'_{\bar{h}}) - \phi(p(\theta'))$$
$$\geq \gamma\ \underline{\lim}\ P_0 P_{\theta'}\ [E|X_\alpha = u]$$
$$\geq \gamma\ \alpha_1 > 0\quad ,$$

where $\alpha_1 = P_0\ \Phi(-\delta f_0(U)\ \sigma^{-1}(U)) > 0$.

Equation (29) demonstrates that a uniform convergence theorem of order better than $O(N^{-1/2})$ is impossible in this discrimination example where (C) is violated.

CHAPTER IV

THE TWO-DECISION COMPOUND TESTING PROBLEM IN THE PRESENCE OF A NUISANCE PARAMETER

1. Introduction.

We now give a formulation of the testing problem considered in Chapter II for testing between the parameter values $\theta = 0$ and 1 in the presence of an unknown nuisance parameter $\tau = (\tau_1,\ldots,\tau_s)$, $s \geq 1$. Let T be a set in R^s with non-empty interior. Let $\mathcal{P}_\theta = \{P_{\theta,\tau} | \tau \varepsilon T\}$ be a family of distributions for $\theta = 0,1$. We shall assume throughout this chapter the existence of a σ-finite measure ν dominating the families $\mathcal{P}_\theta$, $\theta = 0,1$.

Consider now the compound problem of making N decisions, '$\theta_\alpha = 0$ or 1', based on N independent observations X_α, $\alpha = 1,\ldots,N$, where X_α is distributed as $P_{\theta_\alpha,\tau}$ for fixed $\tau \varepsilon T$. Let the loss matrix for the component problem be given by:

$$(1) \qquad \begin{pmatrix} 0 & b \\ a & 0 \end{pmatrix}$$

where $b > 0$ represents loss due to deciding $P_{1,\tau}$ when $P_{0,\tau}$ is the case and $a > 0$ the loss for deciding $P_{0,\tau}$ when $P_{1,\tau}$ is true.

If τ is known, the problem reduces to that of Chapter II. However, we here consider the case where the vector parameter τ is unknown but assumed to be the same for all N problems. We shall give a procedure which first estimates τ and $\bar{\theta}$ based on $X_1,\ldots,X_N$ and then adopts a compound procedure similar to that given by (2.9) in Chapter II with τ replaced by its estimate.

At this point it seems appropriate to remark on the above formulation of the problem. Suppose we temporarily consider the problem where the set T is a finite set of, say, k elements, $\tau_o,\dots,\tau_{k-1}$. Then the problem reduces to that of Chapter III by letting $m = 2k$ with the class $\mathcal{P} = \{P_o,\dots,P_{m-1}\}$ being given by $P_\ell = P_{0,\tau_\ell}$ and $P_{k+\ell} = P_{1,\tau_\ell}$ for $\ell = 0,\dots,k-1$, and by choosing a $2k \times 2$ loss matrix with $L(i,0) = 0$ or a according as $i < k$ or $\geq k$ and $L(i,1) = 0$ or b according as $i \geq k$ or $< k$, where decision j, for $j = 0,1$, corresponds to saying '$\theta = j$'. In fact, in this case τ can vary over T from component problem to component problem. Hence, we arrive at no new problem unless T is at least an infinite set. The selection of T as a set in R^s with non-empty interior will be seen in the proofs to follow. The assumption that τ is the same for all N problems permits the obtaining of estimates of τ which have good asymptotic properties as $N \to \infty$.

With the formulation of T as a set in R^s with non-empty interior and τ the same for all N decision problems, the earlier results do not yield a solution. It is this problem to which we now devote our attention. We will give asymptotic solutions (in the sense of regret risk convergence) in Theorems 7-11.

Before stating the theorems specifically, a few preliminaries are necessary. Let ν be the assumed σ-finite dominating measure for the families $\mathcal{P}_\theta$, $\theta = 0$ or 1. Fix $\tau \in T$, and denote $P_{0,\tau}$, $P_{1,\tau}$ and $P_{\theta,\tau} = X_{\alpha=1}^{\infty} P_{\theta_\alpha,\tau}$ by P_o, P_1, and P_θ respectively. Define

$$(2) \qquad g_i(u) = \frac{dP_i}{d\nu}(u) \qquad \text{for } i = 0,1.$$

Let $\mu = aP_1 + bP_0$ as in Chapter II and note we may proceed exactly as in equations (2.1) - (2.8). We shall suppress τ in all these equations except (2.2), where

(3) $$Z(u,\tau) = bf_0(u) = \{ag_1(u) + bg_0(u)\}^{-1}\, bg_0(u).$$

Consider now a scalar function h and a vector function $k = (k_1,\ldots,k_s)$ such that $h(U)$ is an unbiased estimate of $\theta = 0$ or 1 and $k(U)$ is an unbiased estimate of τ; that is,

(4) $$P_i\, h(U) = i \qquad \text{for } i = 0,1,$$
$$P_i\, k(U) = \tau \qquad \text{for } i = 0,1.$$

By (4), we then form unbiased estimates of $\bar{\theta}$ and τ based on the observations $X_1,\ldots,X_N$ for all N, $\theta \in \Omega_\infty$ by defining the averages

(5) $$\bar{h}(X) = N^{-1} \textstyle\sum_{\alpha=1}^N h(X_\alpha),$$
$$\bar{k}(X) = N^{-1} \textstyle\sum_{\alpha=1}^N k(X_\alpha) \quad .$$

Observe that by (4), $P_\theta \bar{h}(X) = \bar{\theta}$ and $P_\theta \bar{k}(X) = \tau$. For kernel functions h, $k = (k_1,\ldots,k_s)$ such that h, $k_j \in L_2(P_i)$ for $i = 0,1$; $j=1,\ldots,s$, define $\sigma_i^2(h) = P_i(h(U)-i)^2$ and $\sigma_i^2(k_j) = P_i(k_j(U) - \tau_j)^2$. For $p \in [0,1]$, define $\sigma_p^2(k_j) = p\,\sigma_1^2(k_j) + (1-p)\,\sigma_0^2(k_j)$, $j=1,\ldots,s$. Then by Lemma 4 and its analogue for k, we have

(6) $$P_\theta(\bar{h}-\bar{\theta})^2 \leqq \bar{\sigma}^2(h)N^{-1},\ P_\theta\|\bar{k}-\tau\|^2 \leqq C_1^2\, N^{-1}\ ,$$

where $\bar{\sigma}^2(h) = \max_{i=0,1} \{\sigma_i^2(h)\}$ and $C_1^2 = \max_{i=0,1} \sum_{j=1}^s \sigma_i^2(k_j)$.

From the above formulation, it now becomes natural to consider the compound procedure formed by substituting $\bar{h}(X)$ and $\bar{k}(X)$ for $\bar{\theta}$ and τ respectively in the non-randomized simple procedure given by (2.5) with $\delta_{\bar{\theta}} = 0$. However, since $Z(u,\tau)$ is only defined if τ is in T, we must "truncate" $\bar{k}(X)$ to T. Hence, let $\bar{k}^*$ denote a specified truncation of $\bar{k}$ to T such that $\bar{k}^* = \bar{k}$ if $\bar{k} \varepsilon T$ and if $\bar{k} \notin T$, then $\bar{k}^*$ is a point in T within a distance of N^{-1} of a minimizer of $\|\bar{k} - \tau\|$ on the closure of the set T. A constructive method of truncating when T is a convex set is given in Appendix 2.

We are now able to present a well-defined non-simple, non-randomized procedure $t'_{\bar{h},\bar{k}^*} = (t'_{\bar{h},\bar{k}^*}(x_1),\ldots,t'_{\bar{h},\bar{k}^*}(x_N))$ with coordinate functions

(7) $$t'_{\bar{h},\bar{k}^*}(x_\alpha) = 1 \text{ or } 0 \text{ as } Z(x_\alpha,\bar{k}^*) < \text{ or } \geq \bar{h}, \quad \alpha = 1,\ldots,N.$$

The risk of this procedure under P_θ will be denoted by $R(\theta,t'_{\bar{h},\bar{k}^*})$. Under certain regularity assumptions this procedure will be shown to have good, uniform in $\theta \varepsilon \Omega_\infty$, asymptotic properties in the sense of regret risk convergence.

Certain assumptions will be needed in the proofs of Theorems 7-11. Let $\tau \varepsilon T$ be fixed.

Assumption (A_1): There exist functions h and $k = (k_1,\ldots,k_s)$ such that (4) holds and h, $k_j \varepsilon L_3 (P_i)$ for $i = 0,1$; $j = 1,\ldots,s$.

Assumption (A_2): The covariance matrix of $(h,k_1,\ldots,k_s)$ under P_i, denoted by V_i, is of rank $s + 1$, $i = 0,1$.

When they exist, define

(8) $$Z'_j(u,\tau) = \left.\frac{\partial Z}{\partial \tau_j}\right|_\tau$$

and $$Z''_{jk}(u,\tau) = \left.\frac{\partial Z}{\partial \tau_j \partial \tau_k}\right|_\tau \quad .$$

If $s = 1$, denote Z'_1 and Z''_{11} by Z' and Z'' respectively. Also, let

$S_\delta = \{\tau' \varepsilon R^s \mid \|\tau' - \tau\| < \delta\}$.

Assumption (B_1): For some $\delta = \delta(\tau) > 0$ and for almost all $u(\nu)$, $Z(u,\tau)$ admits continuous first-order partial derivatives (relative to T) for non-isolated points $\tau' \varepsilon S \cap T$. Furthermore, there exists a function $M_1 \varepsilon L_1(P_i)$ for $i = 0,1$ such that $|Z_j'(u,\tau')| \leq M_1(u)$ a.e. ν on $S_\delta \cap T$ for $j = 1,\ldots,s$.

Assumption (B_2): For some $\delta = \delta(\tau) > 0$ and for almost all $u(\nu)$, $Z(u,\tau)$ admits continuous second-order partial derivatives (relative to T) for non-isolated points $\tau' \varepsilon S_\delta \cap T$. Furthermore, $P_i|Z_j'(u,\tau)| < \infty$ and there exists a function $M_2 \varepsilon L_1(P_i)$ for $i = 0,1$ such that $|Z''_{jk}(u,\tau')| \leq M_2(u)$ a.e. ν on $S_\delta \cap T$ for $j,k = 1,\ldots,s$.

2. A Convergence Theorem in the Presence of a Nuisance Parameter.

Theorem 7.

Let τ be any interior point of T for which assumptions (A_1), (A_2), and (B_2) hold. Then, $R(\theta, t'_{\bar{h},\bar{k}*}) - \phi(\bar{\theta}) = O(N^{-(1/2)+\varepsilon})$ for $\varepsilon > 0$.

Proof. Since τ is an interior point of T, we assume, without loss of generality, that the δ of assumption (B_2) is such that $S_\delta \subset T$.

Identify $t_\zeta = t'_{\bar{h},\bar{k}*}$ in (1.15) of Lemma 5 to obtain,

$$
\begin{aligned}
(9) \qquad & R(\theta, t'_{\bar{h},\bar{k}*}) \\
& = \{a\bar{\theta}\, P_\theta P_1(1 - t'_{\bar{h},\bar{k}*}(U)) + b(1-\bar{\theta})\, P_\theta P_0 t'_{\bar{h},\bar{k}*}(U)\} \\
& + a\, N^{-1} \textstyle\sum_{\alpha\varepsilon I_1} P_\theta P_1(t'_{\bar{h},\bar{k}*}(U) - t'_{\bar{h}(\alpha),\bar{k}(\alpha)*}(U)) \\
& + b\, N^{-1} \textstyle\sum_{\alpha\varepsilon I_0} P_\theta P_0(t'_{\bar{h}(\alpha),\bar{k}(\alpha)*}(U) - t'_{\bar{h},\bar{k}*}(U)) ,
\end{aligned}
$$

where $I_i = \{\alpha | \theta_\alpha = i\}$, $i = 0,1$. Let the three terms on the right-hand side of (9) be denoted by A_N, B_N, and C_N respectively.

We establish the theorem by showing that: (i) $A_N - \phi(\bar{\theta}) = O(N^{-1/2})$ uniformly in $\theta \in \Omega_\infty$, and (ii) B_N and C_N are of $O(N^{-(1/2)+\varepsilon})$, $\varepsilon > 0$, uniformly in $\theta \in \Omega_\infty$.

(i) Define $A_N' = P_\theta\{a\bar{\theta} P_1(1-t_{\bar{h}}'(U)) + b(1-\bar{\theta}) P_0 t_{\bar{h}}'(U)\}$, where $t_{\bar{h}}'(u) = 1$ or 0 as $Z(u,\tau) <$ or $\geqq \bar{h}$. Then express $A_N - \phi(\bar{\theta})$ as

$$(10) \qquad A_N - \phi(\bar{\theta}) = A_N' - \phi(\bar{\theta}) + A_N - A_N' \ .$$

Observe that with μ replacing μ' in equality (2.12) and part (i) of the proof of Theorem 2, we see by (6) that

$$(11) \qquad A_N' - \phi(\bar{\theta}) \leqq N^{-1/2} (a+b)\, \bar{\sigma}(h).$$

Next consider the term $A_N - A_N'$ in (10). Again by a cancellation argument of the type used to develop (2.12), we may write

$$(12) \qquad A_N - A_N' = P_\theta\, \mu\, \{(\bar{\theta}-Z(u,\tau))\, ([E] - [F])\ ,$$

where $E = \{Z(u,\tau) < \bar{h} \leqq Z(u,\bar{k}^*)\}$ and $F = \{Z(u,\bar{k}^*) < \bar{h} \leqq Z(u,\tau)\}$. Under the $P_\theta \times \mu$ integral subtract and add $Z(u,\bar{k}^*)\, ([E] - [F])$ and bound $(\bar{\theta}-Z(u,\bar{k}^*))\, ([E]-[F])$ by $|\bar{h}-\bar{\theta}|$ and $(Z(u,\bar{k}^*) - Z(u,\tau))\, ([E] - [F])$ by $|Z(u,\bar{k}^*) - Z(u,\tau)|$ to obtain

$$(13) \qquad A_N - A_N' \leqq (a+b)\, P_\theta|\bar{h}-\bar{\theta}| + P_\theta\, \mu|Z(U,\bar{k}^*) - Z(U,\tau)|\ .$$

In the second term on the right-hand side of (13) partition the space under the P_θ integral into $G_\delta = \{||\bar{k}-\tau|| < \delta\}$ and its complement.

By Assumption (B_2) and our choice of δ, expand $Z(u,\bar{k}^*) = Z(u,\bar{k})$ on G_δ about $Z(u,\tau)$ in a second-order Taylor expansion and bound Z_{jk}'' by M_2 to obtain

(14) $$|Z(u,\bar{k}^*) - Z(u,\tau)|$$

$$\leqq |\textstyle\sum_{j=1}^{s}(\bar{k}_j-\tau_j)\, Z_j'(u,\tau)| + \frac{1}{2}\sum_{k,j} |\bar{k}_k-\tau_k|\, |\bar{k}_j-\tau_j|\, M_2(u)$$

$$\leqq \textstyle\sum_{j=1}^{s} |\bar{k}_j-\tau_j|\, |Z_j'(u,\tau)| + \frac{1}{2}\, \|\bar{k}-\tau\|^2\, M_2(u).$$

Use the Schwarz s-space inequality in the $P_\theta \times \mu$ integral of the first term on the right-hand side of (14) and inequality (6) in the $P_\theta \times \mu$ integral of the both terms to obtain

(15) $$P_\theta\mu\, \{|Z(U,\bar{k}^*) - Z(U,\tau)|\, [G_\delta]\}$$

$$\leqq N^{-1/2}\, C_1\, a_1 + \frac{1}{2} N^{-1}\, C_1^2\, \mu\, (M_2(U))\ ,$$

where $a_1 = \left\{\sum_{j=1}^{s} (\mu|Z_j'(U,\tau)|)^2\right\}^{\frac{1}{2}}$.

Since $|Z(u,\bar{k}^*) - Z(u,\tau)| \leqq 1$, we have by Tchebichev's inequality and (6),

(16) $$P_\theta\, \mu\, |Z(U,\bar{k}^*) - Z(U,\tau)|\, (1 - [G_\delta])$$

$$\leqq (a+b)\, \delta^{-2}\, P_\theta\, \|\bar{k}-\tau\|^2$$

$$\leqq (a+b)\, \delta^{-2}\, C_1^2\, N^{-1} \quad .$$

Hence, (15) and (16) together with the Schwarz inequality and (6) used to obtain $P_\theta\, |\bar{h}-\bar{\theta}| \leqq N^{-1/2}\, \bar{\sigma}(h)$ imply by inequality (13) that

(17) $$A_N - A_N' \leqq N^{-1/2} \left\{(a+b)\, \bar{\sigma}(h) + a_1\, C_1\right\} + O(N^{-1})$$

uniformly in $\theta \in \Omega_\infty$.

Substitution of (11) and (17) into (10) completes the proof of (i).

(ii) We shall now bound the term B_N in (9). Without loss of generality we assume $N\bar{\theta} \geq 1$. Let $0 < \varepsilon < \frac{1}{2}$ be given. Fix $\alpha \varepsilon I_1$ and define the sets $E = \{\|\bar{k}-\tau\| \geq N^{-(1/2)(1-\varepsilon)}\}$ and $E_\alpha = \{\|\bar{k}^{(\alpha)}-\tau\| \geq N^{-(1/2)(1-\varepsilon)}\}$. We shall need the following pointwise inequality:

(18) $$t_{\bar{h},\bar{k}^*}(u) - t_{\bar{h},\bar{k}^{(\alpha)*}}(u)$$

$$\leq [Z(u,\bar{k}^*) < \bar{h}]\ [Z(u,\bar{k}^{(\alpha)*}) \geq \bar{h}^{(\alpha)}](1-[E])(1-[E_\alpha])+[E]+[E_\alpha].$$

We now bound the $P_\theta \times P_1$ integral of the right-hand side of (18).

Observe that by a change of variable and an elementary set inclusion, we have

(19) $$P_\theta P_1([E] + [E_\alpha]) = 2\ P_\theta[E]$$

$$\leq 2 \sum_{j=1}^{s} P_\theta[|\bar{k}_j-\tau_j| \geq s^{-1/2}\ N^{-(1/2)(1-\varepsilon)}].$$

By a B-E normal approximation to each of the summands on the right of (19), we have for $j = 1,\ldots,s$,

(20) $$P_\theta[|\bar{k}_j-\tau_j| \geq s^{-1/2}\ N^{-(1/2)(1-\varepsilon)}]$$

$$\leq 2\ \{1 - \Phi(s^{-1/2}\ N^{\varepsilon/2}\ \sigma_{\bar{\theta}}^{-1}(k_j)\}$$

$$+ 2\ \beta\ N^{-1/2}\ b_j(\bar{\theta}) \quad ,$$

where $b_j(\bar{\theta}) = \sigma_{\bar{\theta}}^{-3}(k_j)\{\ \bar{\theta} P_1\ |k_j(U)-\tau_j|^3 + (1-\bar{\theta})\ P_0\ |k_j(U)-\tau_j|^3\}$.

We bound from above the first term in (20) by noting that $1 - \Phi(s^{-1/2}\ N^{\varepsilon/2}\ \sigma_{\bar{\theta}}^{-1}(k_j)) \leq 1 - \Phi(s^{-1/2}\ N^{\varepsilon/2}\ d_j^{-1})$ since for all $\theta \varepsilon \Omega_\infty$ $d_j \equiv \max_{i=0,1}\ \{\sigma_i^2(k_j)\} \geq \sigma_{\bar{\theta}}^2(k_j)$. Then, by the exponential tail inequality $1 - \Phi(x) \leq \Phi'(0)\ x^{-1} \exp\{-\frac{1}{2}x^2\}$, for $x > 0$, (see Feller [3], p. 166), we have

$$(21) \qquad 1 - \Phi(s^{-1/2} N^{\varepsilon/2} \sigma_{\bar{\theta}}^{-1}(k_j))$$
$$\leq \Phi'(0)\, s^{1/2} N^{-\varepsilon/2} d_j \exp\{-(1/2)\, s^{-1} d_j^{-2} N^{\varepsilon}\} \quad .$$

Define $b_j = \max_{p\varepsilon[0,1]} b_j(p)$. The second term on the right-hand side of (20) is then $O(2\beta b_j N^{-1/2})$ uniformly in $\theta \varepsilon \Omega_\infty$, since $b_j(\bar{\theta}) \leq b_j$ for all $\theta \varepsilon \Omega_\infty$. This bound asymptotically dominates the exponential bound in (21) and when it is substituted with (21) into (20) for $j = 1,\ldots,s$ we see that by inequality (19),

$$(22) \qquad P_\theta P_1([E] + [E_\alpha]) = O(b_0 N^{-1/2}) \text{ , uniformly in } \theta \varepsilon \Omega_\infty \text{, where}$$

$b_0 = 4\beta \sum_{j=1}^{s} b_j$.

To bound the $P_\theta \times P_1$ integral of the first term on the right-hand side of (18), choose $N_0 = N_0(\tau)$ sufficiently large such that $N_0^{-(1/2)(1-\varepsilon)} < \delta$. Then, by assumptions ($B_2$), we may on the set $E^c \cap E_\alpha^c$ (c denoting complement) expand $Z(u,\bar{k}^*) = Z(u,\bar{k})$ and $Z(u,\bar{k}^{(\alpha)*}) = Z(u,\bar{k}^{(\alpha)})$ about τ in second-order Taylor expansions and bound them from below and above as follows:

$$(23) \qquad Z(u,\bar{k}) \geq Z(u,\tau) + \sum_{j=1}^{s} (\bar{k}_j - \tau_j)\, Z_j'(u,\tau) - \tfrac{1}{2} N^{-1+\varepsilon} M_2(u)$$

$$\text{and,} \qquad Z(u,\bar{k}^{(\alpha)}) \leq Z(u,\tau) + \sum_{j=1}^{s} (\bar{k}_j^{(\alpha)} - \tau_j)\, Z_j'(u,\tau) + \tfrac{1}{2} N^{-1+\varepsilon} M_2(u).$$

Define, $w(x_\ell) = (h(x_\ell) - \theta_\ell, k_1(x_\ell) - \tau_1, \ldots, k_s(x_\ell) - \tau_\ell)$, for $\ell = 1,\ldots,N$, and $y(u) = (1, -Z_1'(u,\tau), \ldots, -Z_s'(u,\tau))$. Inequalities (23) applied to the first term on the right of (18) together with some algebraic manipulation now imply that this term is bounded from above by the function $[F_\alpha]$, where

$$(24)\qquad [F_\alpha] = [N(Z(u,\tau)-\bar\theta) - \tfrac{1}{2}N^\varepsilon M_2(u) - (y(u),w(x_\alpha) + \textstyle\sum_{\ell\varepsilon I_0} w(x_\ell))$$

$$< \textstyle\sum_{\ell\varepsilon I_1,\ell\neq\alpha}(y(u),w(x_\ell)) \leqq N(Z(u,\tau) - \bar\theta) + \tfrac{1}{2}N^\varepsilon M_2(u)$$

$$- (y(u),\ w(u) + \textstyle\sum_{\ell\varepsilon I_0} w(x_\ell))].$$

Condition the $P_\theta \times P_1$ integral of $[F_\alpha]$ on u, x_α, and x_ℓ, $\ell \varepsilon I_0$ and apply a B-E approximation, obtaining an upper bound on the conditional probability given by

$$(25)\qquad \min\{1,(N\bar\theta-1)^{-1/2}\ (\Phi'(0)\{N^\varepsilon\alpha_1(u) + \alpha_2(u,x_\alpha)\} + 2\beta\alpha_3(u))\}\ ,$$

where $\alpha_1(u) = \sigma_1^{-1}((y(u),w))M_2(u)$, $\alpha_2(u,x_\alpha) = \sigma_1^{-1}((y(u),w))|(y(u),w(u)-w(x_\ell))|$ and $\alpha_3(u) = \sigma_1^{-3}((y(u),w))\ P_1|(y(u),w)|^3$, with $\sigma_1^2(t)$ denoting the variance of t(V) under P_1. Assume for the moment that α_i, i= 1,2,3 are integral with respect to $P_{\theta_\alpha} \times P_1$. Then,

$$(26)\qquad P_\theta P_1[F_\alpha] \leqq \min\{1,(N\bar\theta-1)^{-1/2}\ (N^\varepsilon\alpha_1^* + \alpha_0^*)\}\ ,$$

where $\alpha_1^* = \Phi'(0)P_1\alpha_1(U)$ and $\alpha_0^* = \Phi'(0)P_1P_1\alpha_2(U,V) + 2\beta P_1\alpha_3(U)$.

Recalling the definition of B_N and the fact that the function $[F_\alpha]$ bounds the first term on the right-hand side of (18), we see that equations (22) and (26) substituted into the $P_\theta \times P_1$ integral of inequality (18) and summed over all $\alpha \varepsilon I_1$ imply

$$(27)\qquad B_N \leqq a\ N^{-1/2}\Big\{N^{1/2}\bar\theta\ \min\{1,(N\bar\theta-1)^{-1/2}\ (N^\varepsilon\alpha_1^*+\alpha_0^*)\}\ + O(b_0)\Big\}\ .$$

Inequality (2.14) when substituted into (27) with $C = N^\varepsilon\alpha_1^* + \alpha_0^*$ and $p = \bar\theta$ yields

$$(28)\qquad B_N = O(N^{-(1/2)+\varepsilon})\ \text{uniformly in}\ \theta \varepsilon \Omega_\infty.$$

It must be recalled that equation (28) was derived under the assumption that α_i for i = 1,2,3 were integrable with respect to $P_1 \times P_{\theta_\alpha}$, $\alpha \in I_1$. Observe that

$$(29) \qquad \sigma_1^2((y(u),w)) = y(u)\, V_1\, y'(u) = y(u)\, \Gamma\, D\, \Gamma'\, y'(u),$$

where Γ is an orthogonal matrix, D a diagonal matrix with diagonal elements d_i, $i = 1,\ldots,s+1$. Let $v = y(u)\Gamma$ and $\lambda_1^* = \min_i d_i$. Then,

$$(30) \qquad vDv' \geqq \sum_{i=1}^{s+1} v_i^2\, \lambda_1^* = \|y(u)\|^2\, \lambda_1^*.$$

Therefore, weakening by the Schwarz inequality for s+1 space in the numerators of $\alpha_2(u,x_\alpha)$ and $\alpha_3(u)$ and bounding the denominators from below by (29) and (30), we have

$$(31) \qquad \alpha_2(u,x_\alpha) \leqq \|w(u) - w(x_\alpha)\|\, (\lambda_1^*)^{-1/2}$$

$$\alpha_3(u) \leqq P_1\, \|w\|^3 (\lambda_1^*)^{-3/2} .$$

Also, note that (29) and (30) imply

$$(32) \qquad \alpha_1(u) \leqq M_2(u)\, (\lambda_1^*)^{-1/2}$$

since $\|y(u)\|^2 = 1+\sum_{j=1}^{s}\{Z_j'(u,\tau)\}^2 \geqq 1$. Integrability of α_i for i=1,2,3 now follows from (31) and (32) by observing that $M_2 \in L_1(P_1)$ by assumption (B_2), $\|w\| \in L_3(P_1)$ by assumptions (A_1) and the c_r-inequality (Loève [9], p.155), while assumption (A_2) guarantees $\lambda_1^* > 0$. This completes the proof that $B_N = O(N^{-(1/2)+\varepsilon})$ uniformly in $\theta \in \Omega_\infty$. A similar argument shows that $C_N = O(N^{-(1/2)+\varepsilon})$ uniformly in $\theta \in \Omega_\infty$, and (ii) is proved.

The proof is now established by (i), (ii), and equality (9).

Please note that the order here obtained has the factor $N^{+\varepsilon}$, $0 < \varepsilon < \frac{1}{2}$. We were unable, in general, to remove this factor and obtain convergence rates as good as those of Theorems 2 and 5. However, in two later results (Theorems 10 and 11 below) two interesting and rather revealing cases where this factor can be eliminated are given.

3. Examples for Theorem 7.

Three examples satisfying Theorem 7 are given.

Example 1.

Let T be the subset of R^s given by $T = \{(\tau_1,\ldots,\tau_s) \mid \sum_{j=1}^{s} \tau_j < \frac{1}{2}(1 - (s+1)\gamma),\ \tau_j > 0\}$, where γ is a fixed constant such that $0 < \gamma < (s+1)^{-1}$. Note that T is a non-empty open convex subset of R^s. Fix $\tau \in T$ and let the generic random variable $U = (U_1,\ldots,U_{2s+2})$ have the multinomial distribution for i = 0, 1, $P_i\ \{U = u\} = n!\ (\Pi_{j=1}^{2s+2} u_j!)^{-1}\ \Pi_{j=1}^{s+1}\{(\tau_j + i\gamma)^{u_j}(\tau_j + (1-i)\gamma)^{u_{s+1+j}}\}$, where $\sum_{j=1}^{2s+2} u_j = n$ and $\sum_{j=1}^{s+1} \tau_j = \frac{1}{2}(1 - (s+1)\gamma)$. We show that assumptions (A_1), (A_2) and (B_2) of Theorem 7 are satisfied.

Define the functions,

$$(33) \qquad h(u) = \{\gamma(s+1)\}^{-1}\ (n^{-1} \sum_{j=1}^{s+1} u_j - \tfrac{1}{2}) + \tfrac{1}{2}$$

$$k_j(u) = (2n)^{-1}\ (u_j + u_{s+1+j}) - \tfrac{1}{2}\ \gamma \text{ for } j = 1,\ldots,s.$$

Then, since $P_iU_j = n(\tau_j + i\gamma)$ and $P_iU_{s+1+j} = n(\tau_j + (1-i)\gamma)$, for $i = 0,1$; $j = 1,\ldots,s$, it follows that $P_ih(U) = i$ and $P_ik_j(U) = \tau_j$. Assumption (A_1) now follows from boundedness of $|h(u)|$ and $|k_j(u)|$ for $j = 1,\ldots,s$ by $\frac{1}{2}(\{\gamma(s+1)\}^{-1} + 1)$ and $\frac{1}{2}(1-\gamma)$ respectively. Assumption (A_2) is satisfied since $\{h, k_1,\ldots,k_s\}$ forms a linearly independent set of functions in $L_1(P_i)$ for $i = 0,1$.

To verify conditions (B_2), we first define

$$(34) \qquad \psi(u,\tau) = P_1(u)\,\{P_0(u)\}^{-1} = \Pi_{j=1}^{s+1}\,(1 + \gamma\tau_j^{-1})^{u_j - u_{s+1+j}} \quad .$$

Let ψ'_j and ψ''_{jk} be the first- and second-order partials of ψ with respect to τ_j and τ_j,τ_k respectively. The following relationships then hold:

$$(35) \qquad \psi'_j = \gamma\,\psi\,\zeta_j \quad ,$$

$$\psi''_{jk} = \gamma\,\psi\,(\zeta'_{jk} + \gamma\,\zeta_j\zeta_k) \, ,$$

where $\zeta_j(u,\tau) = \{\tau_{s+1}(\tau_{s+1} + \gamma)\}^{-1}\,(u_{s+1} - u_{2s+2}) - \{\tau_j(\tau_j + \gamma)\}^{-1}\,(u_j - u_{s+1+j})$

and $\zeta'_{jk} = \partial\zeta_j/\partial\tau_k$.

Therefore, by expressing $Z(u,\tau) = (a\psi+b)^{-1}\,b$, differentiating as indicated below and substituting equations (35) in the resulting derivatives, we have for $j,k = 1,\ldots,s$,

$$(36) \qquad Z'_j(u,\tau) = -ab\,\gamma\,\psi\,\zeta_j(a\psi+b)^{-2} ,$$

$$Z''_{jk}(u,\tau) = ab\,\gamma\,\psi\,(a\psi+b)^{-3}\,\{\gamma\,\zeta_j\zeta_k(a\psi-b) - (a\psi+b)\,\zeta'_{jk}\} \quad .$$

Hence, observing that $2\,ab\,\psi(a\psi+b)^{-2} \leqq 1$, we obtain from (36),

$$(37) \qquad |Z'_j(u,\tau)| \leqq \tfrac{1}{2}\,\gamma\,|\zeta_j| \quad ,$$

$$|Z''_{jk}(u,\tau)| \leqq \tfrac{1}{2}\,\gamma\,\{\gamma|\zeta_j\zeta_k| + |\zeta'_{jk}|\} \quad .$$

From the definitions of ζ_j and ζ'_{jk} it is readily seen that

$\zeta'_{jk} = (u_{s+1} - u_{2s+2}) \{\tau_{s+1}(\tau_{s+1} + \gamma)\}^{-2} (2\tau_{s+1} + \gamma) + \delta_{jk}(u_j - u_{s+1+j})$

$\{\tau_j(\tau_j + \gamma)\}^{-2} (2\tau_j + \gamma)$, where δ_{jk} is the Kronecker δ. Hence, since $|u_j - u_{s+1+j}| \leq n$ and $(2\tau_j + \gamma) \leq 1 - s\gamma$, for $j = 1,\ldots,s+1$, we have for $j,k = 1,\ldots,s$,

$$(38) \qquad |\zeta_j| \leq n\,(q(\tau_{s+1}) + q(\tau_j))$$

and

$$|\zeta'_{jk}| \leq n(1-s\gamma)\,(q^2(\tau_{s+1}) + q^2(\tau_j)) \, ,$$

where $q(x) = \{x(x + \gamma)\}^{-1} > 0$ for $x > 0$.

The first inequalities of (37) and (38) yield

$(39) \qquad P_i|Z'_j(u,\tau)| \leq \frac{1}{2}\, n\, \gamma\, \{q(\tau_{s+1}) + q(\tau_j)\} < \infty$. To complete the verification of assumption (B_2) define $\delta = \delta(\tau) = \frac{1}{2}\min \{\tau_1,\ldots,\tau_s, s^{-1/2}\,\tau_{s+1}\}$, where $\sum_{j=1}^{s+1} \tau_j = \frac{1}{2}(1 - (s+1)\gamma)$. Define the hyperplanes $H_j = \{\tau \in R^s |\ \tau_j = 0\}$, $j = 1,\ldots,s$ and $H_{s+1} = \{\tau \in R^s |\ \sum_{j=1}^{s}\tau_j = \frac{1}{2}(1 - (s+1)\gamma)\}$. These s+1 hyperplanes intersected with the closure of T form the boundary of T. The distance between H_j and τ is given by τ_j for $j = 1,\ldots,s$ and by $s^{-1/2}\tau_{s+1}$ for $j = s+1$. Hence $S_\delta \subset T$ since T is convex and the radius δ is half the distance of τ to the closest boundary point of T in the bounding hyperplanes H_j, $j = 1,\ldots,s+1$.

We now define the function M_2. Observe that if $\tau' = (\tau'_1,\ldots,\tau'_s) \in S_\delta$, then $\tau'_j > \frac{1}{2}\tau_j$ for $j = 1,\ldots,s+1$, where $\sum_{j=1}^{s+1} \tau'_j = \frac{1}{2}(1 - (s+1)\gamma)$. Hence, with $q(x)$ a strictly decreasing function on $(0,\infty)$ we have $q(\tau'_j) < q(\frac{1}{2}\tau_j)$ for $j = 1,\ldots,s+1$. Thus, define $M_2(u) = n\,\gamma\, q_0^2\, \{2\, n\, \gamma + (1-s\gamma)\}$, where $q_0 = q_0(\tau) = \max_{j=1,\ldots,s+1} q(\frac{1}{2}\tau_j)$.

Then, by (37) and (38) we see $|Z''_{jk}(u,\tau')| \leqq M_2(u)$ a.e. ν for $j,k = 1,\ldots,s$ if $\tau' \in S_\delta$. This together with (39) completes the verification of (B_2).

We have now shown that assumptions (A_1), (A_2), and (B_2) are met for any $\tau \in T$ and hence Theorem 7 is valid for Example 1 with any fixed τ in T.

We now give two examples in which $s = 1$.

Example 2.

Let U be the generic name for the X_α's. With $s = 1$ and $T = (0,\infty)$, fix $\tau \in T$. The distribution of U under P_i is normal with mean i and variance τ for $i = 0,1$. Represent $k_1(u)$ by $k(u)$ and define

(40) $$h(u) = u \quad , \quad k(u) = u^2 - u.$$

Then, $P_i h(U) = i$ and $P_i k(U) = \tau$ for $i=0,1$, and fixed τ. Hence, assumption (A_1) is satisfied, since all absolute moments $P_i|U|^k$, $k = 1,2,\ldots$ are finite for $i = 0,1$. Assumption (A_2) is satisfied since h and k are linearly independent and non-degenerate in $L_1(P_i)$ for $i = 0,1$.

To see that (B_2) is satisfied, let ν be Lebesgue measure, and note that for $i = 0,1$, $g_i(u) = (2\pi\tau)^{-1/2} \exp\{-(2\tau)^{-1}(u-i)^2\}$.

Hence, (3) implies

(41) $$Z(u,\tau) = b\,\{a \exp \tau^{-1}(u - \tfrac{1}{2}) + b\}^{-1} \quad .$$

We see that Z has first and second continuous partials with respect to τ on $(0,\infty)$ which are given by

(42) $$Z'(u,\tau) = \frac{2ab\ \zeta(u,\tau)}{\tau(a \exp \zeta(u,\tau) + b \exp\{-\zeta(u,\tau)\})^2}$$

and

(43) $$Z''(u,\tau) = \frac{-4ab\ \zeta(u,\tau)}{\tau^2\ (a \exp \zeta(u,\tau) + b \exp\{-\zeta(u,\tau)\})^2}$$

$$+ \frac{4ab\ \zeta^2(u,\tau)\ (a \exp \zeta(u,\tau) - b \exp\{-\zeta(u,t)\})}{\tau^2\ (a \exp \zeta(u,\tau) + b \exp\{-\zeta(u,\tau)\})^3}$$

where $\zeta(u,\tau) = (2\tau)^{-1}\ (u - \frac{1}{2})$.

Observe that for t real, $|t|\{a \exp t+b \exp(-t)\}^{-2} \leqq \frac{1}{2}\max\{a^{-2},b^{-2}\}$ and $t^2\ \{a \exp t + b \exp(-t)\}^{-2} \leqq \frac{1}{2} \max \{a^{-2},b^{-2}\}$. Hence, from (42) and (43) we obtain

(44) $$|Z'(u,\tau)| \leqq \tau^{-1}\ c_o,$$
$$|Z''(u,\tau)| \leqq 4\tau^{-2}\ c_o,$$

where $c_o = \max\ \{a^{-1}b,\ ab^{-1}\}$. The two inequalities of (44) together with $\delta = \frac{1}{2}\tau$ and $M_2(u) = 16\ c_o\tau^{-2}$ imply assumption (B_2). To see this, suppose $\tau' \in S_\delta=((1/2)\tau,(3/2)\tau)$. Then, by (44), $|Z''(u,\tau')| \leqq 4(\tau')^{-2}\ c_o < 16\ c_o\ \tau^{-2}$. Therefore, assumptions (A_1), (A_2), and (B_2) are valid in Example 2 for fixed $\tau \in (0,\infty)$ and hence Theorem 7 holds for such a τ.

Example 3.

In the α^{th} component decision problem, let $X_\alpha = (X_{\alpha 1},\ldots,X_{\alpha n})$, $n \geqq 2$ be n independent random variables, each distributed as normal with mean $\theta_\alpha = 0$ or 1, and variance $\tau \in T = (0,\infty)$. With U as the generic name for the X_α's, define

$$(45) \qquad h(u) = \bar{u} = n^{-1} \sum_{i=1}^{n} u_i \quad ,$$

$$k(u) = (n-1)^{-1} \sum_{i=1}^{n} (u_i - \bar{u})^2 \quad ,$$

where $k(u)$ denotes $k_1(u)$ of Theorem 7.

Then, $h(u)$ and $k(u)$ are unbiased estimates of i and τ under P_i for $i = 0,1$ and fixed τ in $(0,\infty)$. Defining ν as n-dimensional Lebesgue measure, an analysis similar to that of Example 2 shows that conditions (A_1), (A_2) and (B_2) of Theorem 7 are satisfied for such a τ. Hence, Theorem 7 holds for Example 3 with $h(u)$ and $k(u)$ defined by (45).

Please note that in Example 3, we have "bunching" of observations on each component problem; that is, we make n, $n \geq 2$, independent observations for each component problem. This "bunching" is what allows obtaining the stronger result for this example via Theorem 10 below.

4. Uniform Theorems in the Parameter τ .

Two theorems are presented in which convergence of the regret risk function is made uniform in $\tau \in C$ (as well as in $\theta \in \Omega_\infty$), where C is a suitably chosen compact subset of T. Also, it is shown that, in Example 3 of section 4.3, uniformity in τ on $(0,\infty)$ cannot be obtained for a wide class of sequences θ in Ω_∞.

Theorem 8.

Let T be a non-empty open convex set in R^s and let C be a compact subset of T. Assume that (A_1), (A_2), and (B_2) hold for all $\tau \in T$ and for $\tau \in C$; $i = 0,1$; $j = 1,\ldots,s$ we have:

(i) $P_{i,\tau} \; |Z_j'(U,\tau)| \leqq A < \infty$,

(ii) $P_{i,\tau} \; M_2(U) \leqq M_2 < \infty$, where $M_2(u)$ exists by (B_2) ,

(iii) $P_{i,\tau} \; |h(U)|^3 \leqq H < \infty$, $P_{i,\tau} |k_j(U)|^3 \leqq K < \infty$,

(iv) $\lambda^*_{i,\tau} \geqq \lambda^* > 0$, where $\lambda^*_{i,\tau}$ is the minimum eigenvalue of $V_{i,\tau}$.

Then, for $\varepsilon > 0$, $R(\theta, t'_{\bar{h},\bar{k}^*}) - \phi(\bar{\theta}) = O(N^{-(1/2)+\varepsilon})$ uniformly in $\theta \in \Omega_\infty$ and $\tau \in C$.

Proof. Since C is compact and T forms an open covering of C, there exists a $\delta > 0$ such that for every $\tau \in C$, $S_\delta(\tau) \subset T$. With $\delta > 0$, which is now independent of $\tau \in C$, proceed exactly as in the proof of Theorem 7. To complete the proof we need only show that the bounds obtained in the proof of Theorem 7 are uniform in $\tau \in C$.

Assumption (iii) provides uniform upper bounds in (11) and (16), while assumptions (i), (ii), and (iii) yield uniform upper bounds for the two terms on the right-hand side of (15). Next observe that condition (iii) furnishes a uniform upper bound for d_j, $j = 1,\ldots,s$ in (21). Also, (iii) and (iv) assure that b_0 in (22) is uniformly bounded from above on C. Finally, we need show that $P_{1,\tau} \times P_{\theta_\alpha,\tau}$ integrals of α_i, $i = 1,2,3$ are bounded from above on C. By assumption (iv), $\lambda^*_{i,\tau} \geqq \lambda^* > 0$, for $i = 0,1$, $\tau \in C$; while conditions (ii) and (iii) imply, respectively, that $P_{i,\tau} \; M_2(U)$ and $P_{i,\tau}\|w(U)\|^3$ are uniformly bounded from above for $i = 0,1$, $\tau \in C$. Therefore, by applying the norm triangle inequality in the first inequality of (31) and bounding $\lambda^*_{1,\tau}$ from below by λ^* in (31) and (32), we have that the

α_i for i = 1,2,3 have uniformly bounded integrals in C with respect to $P_{1,\tau} \times P_{\theta_\alpha,\tau}$, $\alpha \in I_1$.

Since all bounds in the proof of Theorem 7 (the bounds for the term C_N being similar) have been shown to be independent of $\tau \in C$, the proof is complete.

The conditions (i) - (iv) of Theorem 8 are satisfied by the three examples given after Theorem 7. We shall verify this statement for Example 1 only.

Note that $q(\tau_j)$ for $j = 1,\ldots,s+1$ is a continuous function on T and hence by compactness of C there exists a constant $q_1 = \max_{j=1,\ldots,s+1,\tau\varepsilon C} q(\tau_j)$. Hence with $A = \gamma n\, q_1$ and $M_2 = n\gamma q_1^2 \{2n\gamma + (1-s\gamma)\}$, we have by (37) and (38), $|Z_j'(u,\tau)| \leqq A$ and $|Z_{jk}''(u,\tau)| \leqq M_2$ on C. Thus, (i) and (ii) are satisfied. Assumption (iii) is satisfied by uniform boundedness of h(u) and k(u) given by (33). Assumption (iv) follows since $\lambda_{i,\tau}^* = \min_{\{\|y\|=1\}} yV_{i,\tau}y'$ is a continuous function of τ for i = 0,1. We have thus established that Theorem 8 holds for Example 1. Detailed analysis of Examples 2 and 3 yield the same result.

We now give a theorem which states under what conditions we can obtain uniform convergence of the regret risk function when s = 1 and $T = [t_1,t_2]$, $t_1 < t_2$, is a closed, bounded interval on the real line. We shall here truncate $\bar{k}$ to $\bar{k}^*$ in T, where $\bar{k}^*$ is given by

(46) $\bar{k}^*(X) = t_1,\ \bar{k}(X),$ or t_2 as $\bar{k}(X) < t_1,\ \varepsilon\, [t_1,t_2]$ or $> t_2$.

Theorem 9.

Let $T = [t_1, t_2]$ be a non-empty, closed, bounded interval of the real line. Assume that (A_1), (A_2), and (B_2) hold for $\tau \in T$ and that for $\tau \in T$, $i = 0,1$,

(i) $\quad P_{i,\tau} \; |Z'(U,\tau)|^2 \leqq A < \infty$,

(ii) $\quad P_{i,\tau} \; M_2(U) \leqq M_2 < \infty$, where $M_2(u)$ exists by (B_2),

(iii) $\quad P_{i,\tau}|h(U)|^3 \leqq H < \infty$, $P_{i,\tau}|k(U)|^3 \leqq K < \infty$,

(iv) $\quad \lambda^*_{i,\tau} \geqq \lambda^* > 0$, where $\lambda^*_{i,\tau}$ is the minimum eigenvalue of $V_{i,\tau}$.

Then, for $\varepsilon > 0$, $R(\theta, t'_{\bar{h},\bar{k}^*}) - \phi(\bar{\theta}) = O(N^{-(1/2)+\varepsilon})$ uniformly in $\theta \in \Omega_\infty$ and $\tau \in T$.

Proof. Fix $\tau \in T$. As in the proof of Theorem 7, write $R(\theta, t'_{\bar{h},\bar{k}^*}) = A_N + B_N + C_N$, where A, B_N, and C_n are three terms on the right-hand side of (9).

Observe that a second-order Taylor expansion (relative to T) of $Z(u,\bar{k}^*)$ about $Z(u,\tau)$ implies

(47) $$P_\theta \mu \; |Z(U,\tau) - Z(u,\bar{k}^*)|$$
$$\leqq P_\theta|\bar{k} - \tau|\mu \; |Z'(U\,,\tau)| + \frac{1}{2} P_\theta(\bar{k} - \tau)^2\mu(M_2(U)),$$

since $|\bar{k}^* - \tau| \leqq |\bar{k} - \tau|$. Now express $A_N - \phi(\bar{\theta})$ as in (10) and bound $A'_N - \phi(\bar{\theta})$ and $A_N - A'_N$ by inequalities (11) and (13) respectively. Substitute inequality (47) into the second term on the right-hand side of (13), weaken by the Schwarz inequality and (6) in $P_\theta \; |\bar{h}-\bar{\theta}| \leqq \bar{\sigma}(h)N^{-1/2}$ and $P_\theta|\bar{k} - \tau| \leqq C_1 \; N^{-1/2}$ and substitute the last two inequalities into the first term on the right-hand side of (13) and into (47), respectively. The resulting inequality is

(48) $$A_N - \phi(\bar{\theta}) \leq \{2(a+b)\,\bar{\sigma}(h) + C_1\mu|Z'(U,\tau)|\}N^{-1/2} + \frac{1}{2}C_1^2\,\mu(M_2(U))N^{-1}.$$

Since assumptions (i), (ii) and (iii) provide uniform bounds on T for $\mu|Z'(U,\tau)|$, $\mu(M_2(U))$ and $\bar{\sigma}(h)$ and C_1, respectively, (48) implies

(49) $$A_N - \phi(\bar{\theta}) = O(N^{-1/2}) \text{ uniformly in } \theta \in \Omega_\infty \text{ and } \tau \in T.$$

We now prove that B_N is of $O(N^{-(1/2)+\varepsilon})$, $0 < \varepsilon < \frac{1}{2}$, uniformly in $\theta \in \Omega_\infty$ and $\tau \in T$. We assume without loss of generality that $N\bar{\theta} \geq 1$. Fix $\alpha \in I_1$ and consider inequality (18). To bound the $P_\theta \times P_1$ integral of the first term on the right-hand side of (18) expand $Z(u,\bar{k}^*)$ and $Z(u,\bar{k}^{*(\alpha)})$ on the set $E^c \cap E^c_\alpha$ in a second-order Taylor expansion about τ and note that $|\bar{k}^* - \bar{k}^{*(\alpha)}| \leq N^{-1}\,|k(x_\alpha) - k(u)|$ to obtain,

(50) $$\begin{aligned}
&[Z(u,\bar{k}^*) < h]\,[Z(u,\bar{k}^{*(\alpha)}) \geq \bar{h}^{(\alpha)}]\,(1 - [E])\,(1-[E_\alpha]) \\
&\leq \Big[N(Z(u,\tau) - \bar{\theta}) + N(\bar{k}^* - \tau)\,Z'(u,\tau) - \tfrac{1}{2}N^{\varepsilon}M_2(u) \\
&\quad - (h(x_\alpha) - 1) - \textstyle\sum_{\ell\varepsilon I_o} h(x_\ell) < \sum_{\ell\varepsilon I_1,\ell\neq\alpha}(h(x_\ell) - 1) \\
&\quad \leq N(Z(u,\tau) - \bar{\theta}) + N(\bar{k}^* - \tau)\,Z'(u,\tau) + \tfrac{1}{2}N^{\varepsilon}M_2(u) \\
&\quad - (h(u) - 1) - \textstyle\sum_{\ell\varepsilon I_o} h(x_\ell) + |k(x_\alpha) - k(u)|\,|Z'(u,\tau)|\Big]
\end{aligned}$$

Let $[F_\alpha]$ denote the right-hand side of (50). Partition the $P_\theta \times P_1$ integral of $[F_\alpha]$ into the sets $\{\bar{k} = \bar{k}^*\}$, $\{\bar{k} > \bar{k}^*\}$ and $\{\bar{k} < \bar{k}^*\}$. On the set $\{\bar{k} = \bar{k}^*\}$, write $\bar{k}^* = \bar{k}$ in $[F_\alpha]$ and enlarge the domain of integration by taking $[\bar{k} = \bar{k}^*] \leq 1$. With $y(u) = (1,-Z(u,\tau))$ and $w(X_\ell) = (h(X_\ell)-1\,,\ k(X_\ell)-\tau)$ apply the B-E normal approximation to the sum of the $N\bar{\theta} - 1$ random variables $(y(u),\ w(X_\ell))$, $\ell \in I_1$, $\ell \neq \alpha$, conditionally on u, x_α, x_ℓ, $\ell \in I_o$, as in developing (25) and (26).

The resulting upper bound for $P_\theta P_1[F_\alpha][\bar{k} = \bar{k}^*]$ is then given by the bound in (26) where the second term of the minimization is increased by the term

(51) $$(N\bar{\theta}-1)^{-1/2}\Phi'(0)P_1 P_{\theta_\alpha}|k(X_\alpha)-k(U))|\ |Z'(U,\tau)|\ \sigma_1^{-1}((y(U),w)),$$

where the P_1 integral is on U, the P_{θ_α} integral on X_α and $\sigma_1^2((y(u),w))$ is for each u, the variance of $(y(u),w(V))$ under P_1 on V. Inequalities (29) and (30) imply that the term (51) is $\leq (N\bar{\theta}-1)^{-1/2}\Phi'(0)(\lambda^*_{1,\tau})^{-1/2}$ $P_1 P_{\theta_\alpha}|k(X_\alpha)-k(U)|$. Since $\lambda^*_{i,\tau}$ is uniformly bounded from below by assumption (iv) and since $P_1\ \|w(U)\|^3$ and $P_1|k(U)|$ and $P_1 M_2(U)$ are uniformly bounded from above by assumptions (iii) and (ii) respectively, this inequality together with (31) and (32) substituted into (26) is seen to yield

(52) $$P_\theta P_1[F_\alpha]\ [\bar{k} = \bar{k}^*] = O(\min\ \{1,(N\bar{\theta}-1)^{-1/2}\ N^\varepsilon\})$$

uniformly in $\tau \in T$.

On the set $\{\bar{k} < \bar{k}^*\}$, write $\bar{k}^* = t_1$ in $[F_\alpha]$ and enlarge the domain of integration under the $P_\theta \times P_1$ integral by taking $[\bar{k} < \bar{k}^*] \leq 1$. Then apply the B-E normal approximation theorem to the sum of the $(N\bar{\theta}-1)$ random variables $h(X_\ell) - 1$, $\ell \neq \alpha$, $\ell \in I_1$, conditionally on u, x_α, x_ℓ, $\ell \in I_o$ to obtain

(53) $$P_\theta P_1[F_\alpha]\ [\bar{k} < \bar{k}^*]$$
$$\leq \min \Big\{1,(N\bar{\theta}-1)^{-1/2}\ [\sigma_1^{-1}(h)\ \Phi'(0)\ \{N^\varepsilon P_1 M_2(U)$$
$$+ P_1 P_{\theta_\alpha}(|h(X_\alpha) - h(U)| + |k(X_\alpha) - k(U)|\ |Z'(U,\tau)|)\}$$
$$+ \sigma_1^{-3}(h)\ P_1\ |h(U) - 1|^3]\Big\}\ .$$

The Schwarz integral inequality applied twice in the $P_1 \times P_{\theta_\alpha}$ term, together with uniformly bounding all terms of (53) in accord with assumptions (i), (ii), (iii) and (iv), yields the result

$$(54) \qquad P_\theta P_1 \, [F_\alpha] \, [\bar{k} < \bar{k}^*] = O(\min \{1, (N\bar{\theta}-1)^{-1/2} N^\varepsilon\})$$

uniformly in $\tau \in T$. A similar analysis shows that

$$(55) \qquad P_\theta P_1 [F_\alpha] \, [\bar{k} > \bar{k}^*] = O(\min\{1, (N\bar{\theta}-1)^{-1/2} N^\varepsilon\})$$

uniformly in $\tau \in T$.

Observe that assumptions (iii) and (iv) imply that $b = \sup_{\tau \in T} b_0(\tau) < \infty$, where $b_0 = b_0(\tau)$ is the bound in (22), while (iii) implies $d = \sup_{\tau \in T} d_1(\tau) < \infty$, for $d_1(\tau)$ in (21). Hence inequalities (18), (22), (50), (52), (54), and (55) now combine to yield

$$(56) \qquad P_\theta P_1 (t_{\bar{h},\bar{k}^*}(U) - t_{\bar{h},\bar{k}}(\alpha)^*(U)) = O(\min \{1, (N\bar{\theta}-1)^{-1/2} N^\varepsilon\})$$

uniformly in $\tau \in T$.

Finally, summing (56) over for all $\alpha \in I_1$ and recalling the definition of B_N, we see that inequality (2.14) with $C = N^\varepsilon$ implies $B_N = O(N^{-(1/2)+\varepsilon})$ uniformly in $\theta \in \Omega_\infty$ and $\tau \in T$. The same is true for the term C_N. Hence, these results for B_N and C_N, (49) and (9) now complete the proof.

An example will now be given to illustrate the distinction between Theorem 7 and uniform results on compact sets as given in Theorem 8 and Theorem 9. Specifically, we will use Example 3 of section 4.3 and show that for this example we can choose a sequence

$\tau_N \to \infty$ such that the sequence of regret risk functions $R_N(\theta, t'_{\bar{h},\bar{k}})$ - $\phi_N(\bar{\theta}) \to \zeta(\xi) > 0$ as $N \to \infty$ for all $\theta \in \Omega_\infty$ such that $\bar{\theta} \to \xi$, $\xi \neq (a+b)^{-1}$ b as $N \to \infty$. Hence, for this example uniformity in τ on the non-compact set $T = (0,\infty)$ is impossible.

Let $\tau_N = N^{1+\delta}$ for some $\delta > 0$ and let $\theta \in \Omega_\infty$ be such that $\bar{\theta} \to \xi$, $\xi \neq (a+b)^{-1}b$ as $N \to \infty$. Observe that for Example 3 of 4.3,

$\phi_N(\bar{\theta}) = a\,\bar{\theta}\,P_1[q(\bar{\theta}) \geq \tau_N^{-1}\, n(\bar{U} - \frac{1}{2})] + b(1-\bar{\theta})\; P_0[q(\bar{\theta}) < \tau_N^{-1}\, n(\bar{U} - \frac{1}{2})]$,

where $q(\bar{\theta}) = \log\{(a\bar{\theta})^{-1}\, b(1-\bar{\theta})\}$. Hence, since $\tau_N^{-1/2}\, n^{1/2}(\bar{U}-i)$ is $N(0,1)$ under P_i for $i = 0,1$, we have

$$(57)\qquad \phi_N(\bar{\theta}) = a\,\bar{\theta}\{\Phi((\tau_N\, n^{-1})^{1/2}\, q(\bar{\theta}) - \tfrac{1}{2}(n\tau_N^{-1})^{1/2})\}$$
$$+ b(1-\bar{\theta})\,\{1-\Phi((\tau_N\, n^{-1})^{1/2}\, q(\bar{\theta}) + \tfrac{1}{2}(n\tau_N^{-1})^{1/2})\}\quad .$$

Noting that $q(\xi) <$ or > 0 according as $\xi >$ or $< (a+b)^{-1}b$, we see that with $\tau_N = N^{1+\delta}$ and $\bar{\theta} \to \xi$, $\xi \neq (a+b)^{-1}b$ as $N \to \infty$, equation (57) implies

$$(58)\qquad \phi_N(\bar{\theta}) \to b(1-\xi) \text{ or } a\xi \text{ according as } \xi > \text{ or } < (a+b)^{-1}b.$$

We now examine $R_N(\theta, t'_{\bar{h},\bar{k}})$ as $N \to \infty$. Observe that for Example 3 $\bar{h}(X_\ell) = \bar{X}_\ell = n^{-1}\sum_{j=1}^{n} X_\ell$, and hence,

$$(59)\qquad R_N(\theta, t'_{\bar{h},\bar{k}}) = aN^{-1}\sum_{\alpha\in I_1} P_\theta[NZ(X_\alpha,\bar{k}) - \bar{X}_\alpha \geq \sum_{\ell\neq\alpha}\bar{X}_\ell]$$
$$+ bN^{-1}\sum_{\alpha\in I_0} P_\theta[NZ(X_\alpha,\bar{k}) - \bar{X}_\alpha < \sum_{\ell\neq\alpha}\bar{X}_\ell]\ ,$$

where $Z(X_\alpha,\bar{k}) = b(a\exp\{n\bar{k}^{-1}(\bar{X}_\alpha - \frac{1}{2})\} + b)^{-1}$. Observe that $(n(N-1)^{-1}\,\tau_N^{-1})^{1/2}\sum_{\ell\neq\alpha}(\bar{X}_\ell - \theta_\ell)$ is $N(0,1)$ under P_θ and is independent of $\bar{X}_\alpha, k(X_1),\ldots,k(X_N)$, where $k(X_\ell) = (n-1)^{-1}\sum_{j=1}^{n}(X_{\ell j} - \bar{X}_\ell)^2$ from (45). Hence, we may integrate with respect to the joint marginal distribution of the N-1 variables $\bar{X}_\ell$, $\ell \neq \alpha$ in each of the summands of (59) to obtain,

(60) $R_N(\theta, t'_{\bar{h},\bar{k}})$

$$= aN^{-1} \sum_{\alpha \epsilon I_1} P_\theta\left\{\Phi(\{n(N-1)^{-1}\tau_N^{-1}\}^{1/2} \{NZ(X_\alpha,\bar{k}) - N\bar{\theta} - \bar{X}_\alpha + \theta_\alpha\})\right\} ,$$

$$+ bN^{-1} \sum_{\alpha \epsilon I_0} P_\theta \left\{1 - \Phi (\{n(N-1)^{-1}\tau_N^{-1}\}^{1/2} \{NZ(X_\alpha,\bar{k}) - N\bar{\theta} - \bar{X}_\alpha + \theta_\alpha\})\right\} .$$

Observe that by our choice of $\tau_N = N^{1+\delta}$, $\delta > 0$, we can conclude that since $|Z(X_\alpha,\bar{k}) - \bar{\theta}| \leq 1$, the variable $\{n(N-1)^{-1}\tau_N^{-1}\}^{1/2} \{NZ(X_\alpha,\bar{k}) - N\bar{\theta}\} \to 0$ in probability as $N \to \infty$, for each $\alpha = 1,\ldots,N$. Also, since $(n\tau_N^{-1})^{1/2} (\bar{X}_\alpha - \theta_\alpha)$ is $N(0,1)$, we have $(n(N-1)^{-1}\tau_N^{-1})^{1/2} (\bar{X}_\alpha - \theta_\alpha) \to 0$ in probability for $\alpha = 1,\ldots,N$. Hence, the sum of these two variables given by the variable

(61) $$\{n(N-1)^{-1}\tau_N^{-1}\}^{1/2} \{NZ(X_\alpha,\bar{k}) - N\bar{\theta} - \bar{X}_\alpha + \theta_\alpha\} \to 0 \text{ in probability}$$

for each $= 1,\ldots,N$.

We now use (61) to obtain a limiting value for (60). Since a continuous function of a random variable converging in probability to a constant converges in probability to the corresponding functional value of that constant, we see from (61), continuity of Φ, the bounded convergence theorem, and the Toeplitz Lemma (see Loève, [9], p. 238) that the limiting value of (60) is given by

(62) $\lim_{N\to\infty} R_N(\theta, t'_{h,\bar{k}})$

$$= a\xi\Phi(0) + b(1-\xi)\Phi(0) = \frac{1}{2}\{a\xi + b(1-\xi)\} .$$

Equations (58) and (62) yield as a limit for the regret risk function the expression

$$(63) \qquad \lim_{N\to\infty}\{R_N(\theta, t'_{\bar{h},\bar{k}}) - \phi_N(\bar{\theta})\} = \tfrac{1}{2}(a+b)\ |\xi-(a+b)^{-1}b| = \zeta(\xi) > 0,$$

where $\bar{\theta} \to \xi \neq (a+b)^{-1}b$ and $\tau_N = N^{1+\delta}$, $\delta > 0$.

This completes the example which shows that uniformity in $\theta \varepsilon \Omega_\infty$ and $\tau \varepsilon T$ is unobtainable for Example 3 where $T = (0,\infty)$ is non-compact. That this is truly a contradiction to regret risk convergence uniform in both $\theta \varepsilon \Omega_\infty$ and $\tau \varepsilon T$ follows from the observation: If uniformity held on both Ω_∞ and T, then for the diagonal sequence $(\bar{\theta}_N, \tau_N)$, $N = 1,2,\ldots$, we would have $R_N(\theta, t'_{\bar{h},\bar{k}}) - \phi_N(\bar{\theta}) \to 0$, which is contradicted by (63) for Example 3 of section 4.3.

5. Specific Results when s = 1.

Let s = 1 and T be an open interval of the real line. Denote τ_1 by τ and k_1 by k, and fix $\tau \varepsilon T$. We give two cases in which the factor $N^{+\varepsilon}$ can be eliminated in the convergence rate of Theorem 7.

Theorem 10.

Let (A_1), (A_2), and (B_1) hold. If $M_1 \varepsilon L_2(P_i)$ and if h and k are independent under P_i for $i = 0,1$, then $R(\theta, t'_{\bar{h},\bar{k}*}) - \phi(\bar{\theta}) = O(N^{-1/2})$ uniformly in $\theta \varepsilon \Omega_\infty$.

Proof. Choose $\delta > 0$ such that $S_\delta \subset T$ and express $R(\theta, t'_{\bar{h},\bar{k}*}) = A_N + B_N + C_N$ as in Theorem 7. Observe that $A_N - \phi(\bar{\theta}) = O(N^{-1/2})$ uniformly in $\theta \varepsilon \Omega_\infty$ as in Theorem 7 with a first-order Taylor expansion in (14).

To obtain a bound for B_N, assume $N\bar{\theta} \geq 1$, fix $\alpha \varepsilon I_1$, and note that

$$(64) \qquad t'_{\bar{h},\bar{k}^*}(u) - t'_{\bar{h}^{(\alpha)},\bar{k}^{(\alpha)*}}(u)$$

$$\leqq [NZ(u,\bar{k}^*) - h(x_\alpha) < \textstyle\sum_{\ell\neq\alpha} h(x_\ell) \leqq NZ(u,\bar{k}^{(\alpha)*}) - h(u)].$$

Let $[F_\alpha]$ denote the right-hand side of (64). If we condition on u, x_α, x_ℓ, $\ell \in I_o$ and $k(x_\ell)$, $\ell = 1,\ldots,N$ in the $P_\theta \times P_1$ integral of $[F_\alpha]$, then the B-E theorem yields, by independence of h and k, a bound for this conditional probability given by

$$(65) \quad (N\bar{\theta}-1)^{-1/2}\left\{\Phi'(0)\{\sigma_1(h)\}^{-1}\{|h(u)-h(x_\alpha)|+N|Z(u,\bar{k}^*)-Z(u,\bar{k}^{(\alpha)*})|\}+b_1\right\},$$

where $b_1 = 2\beta\{\sigma_1(h)\}^{-3} P_1|h(U) - 1|^3$.

In the second term on the right—hand side of (65) expand $Z(u,\bar{k}^*)$ about $Z(u,\bar{k}^{(\alpha)*})$ in a first-order Taylor expansion on $E = \{|\bar{k}^* - \tau| < \frac{1}{2}\delta\} \cap \{|\bar{k}^{(\alpha)*} - \tau| < \frac{1}{2}\delta\}$ to obtain $N|Z(u,\bar{k}^*) - Z(u,\bar{k}^{(\alpha)*})| \leqq |k(u) - k(x_\alpha)|M_1(u)$. On the complement of E bound $|Z(u,\bar{k}^*) - Z(u,\bar{k}^{(\alpha)*})|$ by unity and note that a change of variable, Tchebichev's inequality and (6) imply $P_\theta P_1(1 - [E]) \leqq 2P_\theta[|k^* - \tau| \geqq \frac{1}{2}\delta] \leqq 8\delta^{-2} P_\theta(\bar{k}-\tau)^2 \leqq 8\delta^{-2} C_1^{\,2} N^{-1}$. Hence,

$$(66) \quad NP_\theta P_1|Z(U,\bar{k}^*)-Z(U,\bar{k}^{(\alpha)*})| \leqq P_{\theta_\alpha} P_1|k(U)-k(X_\alpha)|M_1(U)+8\delta^{-2}C_1^2.$$

Finally, weakening by the Schwarz inequality to obtain $P_{\theta_\alpha} P_1|k(U) - k(X_\alpha)|M_1(U) \leqq \{2\, P_1 M_1^{\,2}(U)\}^{1/2}\ \sigma_1(k) = b_2$ and $P_{\theta_\alpha} P_1|h(U) - h(X_\alpha)| \leqq 2^{1/2}\sigma_1(h)$, inequalities (65) and (66) imply

$$(67) \qquad P_\theta P_1[F_\alpha] \leqq \min\left\{1,(N\bar{\theta}-1)^{-1/2}\, b_3\right\}$$

where $b_3 = \Phi'(0)\left\{2^{1/2} + \sigma_1^{-1}(h)(b_2 + 8\sigma^{-2}C_1^2)\right\} + b_1 < \infty$.

Recalling the definition of B_N and summing the $P_\theta \times P_1$ integral of inequality (64) for all $\alpha \in I_1$, we have, by inequality (67) and (2.14) with $C = b_3$ and $p = \bar{\theta}$,

(68) $$N^{1/2} B_N \leq a(1+b_3^2)^{1/2} .$$

Hence, by (68), $B_N = O(N^{-1/2})$ uniformly in $\theta \in \Omega_\infty$. A similar argument holds for C_N, and Theorem 10 is proved.

Note that in Example 3 following Theorem 7 the selection of n, $n \geq 2$, independent observations per problem furnish estimates h and k, given by (45), satisfying the independence condition of Theorem 10.

Theorem 11.

Let (B_1) hold and assume there exists a function $k \in L_3(P_i)$ satisfying (4) such that $\sigma_i^2(k) > 0$ for $i = 0,1$. For almost all $u(\nu)$, let $Z(u,\tau)$ be a strictly monotone function on T. Then, the regret risk function $R(\theta, t'_{\bar{\theta},\bar{k}*}) - \phi(\bar{\theta}) = O(N^{-1/2})$ uniformly in $\theta \in \Omega_\infty$.

Proof. Choose δ of assumption (B_1) such that $S_\delta \subset T$. Identify $t_\zeta = t'_{\bar{\theta},\bar{k}*}$ in Lemma 5 to obtain

(69) $$\begin{aligned} &R(\theta, t'_{\bar{\theta},\bar{k}*}) \\ &= \{a\bar{\theta}P_\theta P_1(1-t'_{\bar{\theta},\bar{k}*}(U)) + b(1-\bar{\theta})P_\theta P_0 t'_{\bar{\theta},\bar{k}*}(U)\} \\ &\quad + a N^{-1} \textstyle\sum_{\alpha\in I_1} P_\theta P_1(t'_{\bar{\theta},\bar{k}*}(U) - t'_{\bar{\theta},\bar{k}(\alpha)*}(U)) \\ &\quad + b N^{-1} \textstyle\sum_{\alpha\in I_0} P_\theta P_0(t'_{\bar{\theta},\bar{k}(\alpha)*}(U) - t'_{\bar{\theta},\bar{k}*}(U)). \end{aligned}$$

Let A_N^*, B_N, and C_N denote the three terms on the right-hand side of (69).

Note that here $A_N^* - \phi(\bar{\theta})$ is equal to the $A_N - A_N'$ term in the proof of Theorem 7 with $\bar{h}$ replaced by $\bar{\theta}$. Hence, replacing $\bar{h}$ by $\bar{\theta}$ in (12) and (13) in the proof of Theorem 7, we obtain

$$(70) \quad A_N^* - \phi(\bar{\theta}) \leqq P_\theta \mu |Z(U,\bar{k}^*) - Z(U,\tau)| .$$

In (70) partition the space under the P_θ integral into $D_\delta = \{|\bar{k}^* - \tau| < \delta\}$ and its complement. For fixed u, expand $Z(u,\bar{k}^*)$ about $Z(u,\tau)$ on D_δ in a first-order Taylor expansion to obtain $P_\theta|Z(u,\bar{k}^*) - Z(u,\tau)| \leqq P_\theta|\bar{k}-\tau|M_1(u) \leqq N^{-1/2}\, C_1 M_1(u)$, where the last inequality follows from the Schwarz integral inequality and (6). Bound $|Z(u,\bar{k}^*) - Z(u,\tau)|$ by unity on the complement of D_δ and note that Tchebichev's inequality and (6) imply $P_\theta(1-[D_\delta]) \leqq \delta^{-2}\, C_1^{\,2}\, N^{-1}$. Hence, from (70) we obtain,

$$(71) \qquad A_N^* - \phi(\bar{\theta}) \leqq N^{-1/2}\, C_1 \mu(M_1(U)) + \delta^{-2}\, C_1^{\,2}\, N^{-1}.$$

To bound the term B_N, assume $N\bar{\theta} \geqq 1$ and fix $\alpha \in I_1$. The monotonicity assumption on Z implies that a unique inverse function of $Z(u,\cdot)$, denoted of Z_u^{-1}, exists on the range of $Z(u,\cdot)$ for almost all $u(\nu)$. Hence,

$$(72) \qquad t'_{\bar{\theta},\bar{k}*}(u) - t'_{\bar{\theta},\bar{k}(\alpha)*}(u) \leqq [F_\alpha],$$

where $F_\alpha = \{\bar{k} < Z_u^{-1}(\bar{\theta}) \leqq \bar{k}^{(\alpha)}\}$ or $\{\bar{k} > Z_u^{-1}(\bar{\theta}) \geqq \bar{k}^{(\alpha)}\}$ according as $Z(u,\cdot)$ is strictly increasing or decreasing on T.

For fixed u, x_α, x_ℓ, $\ell \in I_0$, the sum $\sum_{\ell \in I_1 \ell \neq \alpha}(k(x_\ell) - \tau)$ in F_α falls into an interval of length $|k(x_\alpha) - k(u)|$. Hence a B-E approximation applied to the $P_\theta \times P_1$ of $[F_\alpha]$ conditionally on u, x_α, and

x_ℓ, $\ell \in I_0$, together with weakening the resulting bound by $P_{\theta_\alpha} P_1 |k(x_\alpha) - k(U)| \leq 2^{1/2} \sigma_1(k)$, yields

(73) $$P_\theta P_1[F_\alpha] \leq \min \{1, (N\bar{\theta}-1)^{-1/2} C\} ,$$

where $C = 2^{1/2} \Phi'(0) + 2\beta\{\sigma_1(k)\}^{-3} P_1|k(U)-\tau|^3$.

Hence, recalling the definition of B_N and summing the $P_\theta \times P_1$ integral of inequality (72) for all $\alpha \in I_1$, (73) and (2.14) imply

(74) $$N^{1/2} B_N \leq a(1+C^2)^{1/2} \text{ for all } \theta \in \Omega_\infty .$$

A similar result holds for C_N, which together with (69), (71), and (74) completes the proof.

It is interesting to note that Theorem 11 combines with Theorem 2 of Chapter II to state that if $\bar{\theta}$ <u>or</u> τ is known for the 2 x 2 compound testing problem, then under suitable assumptions (see Theorem 2 and Theorem 11) a regret risk convergence of order $O(N^{-1/2})$ uniformly in $\theta \in \Omega_\infty$ can be obtained. However, the convergence rate in Theorem 7 has an additional factor of N^ε, $\varepsilon > 0$, when <u>both</u> $\bar{\theta}$ and τ are unknown and need to be estimated. Attempts to remove the factor $N^{+\varepsilon}$ when both $\bar{\theta}$ and τ are unknown were unsuccessful except in Theorem 10.

SUMMARY

This thesis has demonstrated that compound decision procedures which are asymptotically optimal in the sense of regret risk convergence are obtainable for a variety of compound decision problems. The existence of such procedures was heuristically argued by Robbins in [10] and substantiated in the compound testing problem for two distributions by Hannan and Robbins in [7]. Motivated by these two papers, we proved convergence theorems for the regret risk function of non-simple, non-randomized procedures which are "Bayes" against estimates $\bar{h}$ of the empirical distribution on Ω. The existence and structure of the estimates $\bar{h}$ are given by Theorem 1 and (1.11).

Three cases were considered: (i) the compound testing problem between two specified distributions; (ii) the general m x n compound decision problem; and (iii) the compound testing problem between two specified families of distributions indexed by a common nuisance parameter.

Theorems 2, 5, and 7 give the basic regret risk convergence theorems for the three respective cases. Theorem 5 is of particular interest since it treats the original problem of Hannan and Robbins (Theorem 4, [7]) in the general m x n compound decision problem. Theorems 2 and 5 have uniform (in $\theta \in \Omega_\infty$) convergence rates of $O(N^{-1/2})$, while Theorem 7 has the slightly slower rate of $O(N^{-(1/2)+\varepsilon})$, $\varepsilon > 0$, caused by added estimation of the nuisance parameter. With the nuisance parameter in an open interval of the real line, removal of the factor $N^{+\varepsilon}$ is established in Theorem 10, if $\bar{h}$

is independent of the estimate of the nuisance parameter, and in Theorem 11, if the empirical distribution on $\Omega = \{0,1\}$ is known.

Theorems 3 and 4 reveal that, in the compound testing problem for two distributions, uniform convergence rates of $o(N^{-1/2})$ and $O(N^{-1})$ are attained if appropriate continuity conditions are imposed on P_0 and P_1. Note that Theorem 4 states conditions under which the procedure (2.9) has, regardless of the size of N, a sum of expected losses for the N problems within a uniform constant of the minimum expected sum of losses among all simple procedures. Theorem 6 generalizes the result of Theorem 4 under a suitable condition on the m x n loss matrix.

Examples illustrating the extent, applicability, and non-vacuity of the sufficient conditions were given for all theorems. Examples were also presented to show that Theorem 6 is false without condition (C) and to demonstrate that uniformity in both the nuisance parameter τ and $\theta \varepsilon \Omega_\infty$ is impossible in Theorem 7.

Finally, we point out that Theorems 2 - 11 can be extended to include the non-simple, randomized procedure which is attained by substituting $\bar{h}$ for $p(\theta)$ (and $\bar{k}$ for τ in Theorems 7 - 11) in the simple randomized procedure which assigns equal probabilities of selection among the columns minimizing $(p(\theta), L^{\nu}f)$ in (1.7). This randomized rule and the proof of this statement are given in Appendix 3.

APPENDIX 1.

Proof that Condition (II") Implies Condition (II') when $\mu = P_{\cdot}$.

See Chapter III for the discussion of conditions II' and II".

Lemma 1.1.

Let X_i, $i = 0,\dots,m-1$ be independent and identically distributed uniform random variables on $[0,1]$. If $0 < k \leqq 1$ and if $Z_i = X_i(1,X)^{-1}$, $i = 1,\dots,m-1$ and $Z = (Z_1,\dots,Z_{m-1})$, then the conditional distribution of Z given $(1,X) = \sum_{i=0}^{m-1} X_i = k$ is uniform on $S = \{z = (z_1,\dots,z_{m-1}) \mid z_i \geqq 0,$ $0 < (1,z) \leqq 1\}$.

Proof. Fix z_i, $i = 1,\dots,m-1$ such that $z_i \geq 0$, $0 < \sum_{i=1}^{m-1} z_i \leqq 1$, Then,

(1) $$P\{(1,X) < k,\ Z_i < z_i,\ i = 1,\dots,m-1\}$$

$$= \int_0^1 \cdots \int_0^1 [(1,x) < k,\ x_i < (1,x)z_i,\ i=1,\dots,m-1]\, dx_0 \cdots dx_{m-1}$$

$$= \left(\prod_{i=1}^{m-1} \int_0^{z_i} d\, y_i\right) \int_0^k y_0^{m-1}\, dy_0 = (\Pi_{i=1}^{m-1} z_i)\, m^{-1}\, k^m,$$

where the second equality follows from the transformation $y_i = x_i(1,x)^{-1}$, $i = 1,\dots,m-1$, $y_0 = (1,x)$, having Jacobian y_0^{m-1} .

Similarly, the marginal distribution of $Y_0 = (1,X)$ is given by

(2) $$P\{Y_0 < k\} = \int_0^1 \cdots \int_0^1 [0 \leq (1,x) < k]\, dx_0 \cdots dx_{m-1} = (m!)^{-1}\, k^m,$$

which follows from the transformation $y_j = \sum_{i=j}^{m-1} x_i$, $j = 0,\dots,m-1$ having unit Jacobian. The lemma follows from (1) and (2) by expressing the conditional density of $(Z_1,\dots,Z_{m-1})$ as the joint density of $(Z_1,\dots,Z_m,(1,X))$ which by (1) equals k^{m-1} divided by the density of $(1,X)$ which by (2) equals $\{(m-1)!\}^{-1}\, k^{m-1}$.

Lemma 1.2.

Let $\mu = P_{.} = \sum_{i=0}^{m-1} P_i$ and let $P_i \bar{Z}^{-1}$ represent the induced distribution on S under the transformation $\bar{Z}$: $u \to Z_1(u),\ldots,Z_{m-1}(u)$, with $Z_i = dP_i/dP_{.}$, for $i = 1,\ldots,m-1$. If for some K", $P_i\bar{Z}^{-1} \leqq K''\lambda_{m-1}$, then there exists a K' such that $P_i f^{-1}[B_j] \leqq K'\lambda_m[B_j]$ for $B_j(v,a,b)$ of the form (3.14) with $K = 1$ and $v_j(b-K) = 0$, where $f = (Z_0,\ldots,Z_{m-1})$.

Proof. Note that by the definition of $P_i f^{-1}$, $P_i\bar{Z}^{-1}$, and the assumption of this lemma, we have for $j = 1,\ldots,m-1$,

$$(3)\qquad P_i f^{-1}[B_j] \leqq P_i\bar{Z}^{-1}([-v_0 \leqq \textstyle\sum_{\ell=1}^{m-1} (v_\ell - v_0)\, z_\ell \leqq a - v_0][z_j \leqq b])$$

$$\leqq K''\lambda_{m-1}\Big([-v_0 \leqq \textstyle\sum_{\ell=1}^{m-1}(v_\ell - v_0)z_\ell \leqq a-v_0][z_j \leqq b][z \in S]\Big)\,.$$

If $j = 0$, replace the second factor in (3) by $[1-b \leqq \sum_{j=1}^{m-1} z_j]$. With $\alpha_{m-1} = \lambda_{m-1}[S]$, we see that the measure $\lambda^*_{m-1} = \alpha^{-1}_{m-1}\,\lambda_{m-1}$, when restricted to S, is uniform on S. Hence, by Lemma 1.1, the right-hand side of (3) equals for $j = 0,\ldots,m-1$,

$$(4)\quad K'\int_0^1\cdots\int_0^1 [0 \leqq (v,x) \leqq a(1,x)][x_j \leqq (1,x)b][0 < (1,x) \leqq 1]dx_0\ldots dx_{m-1},$$

where $K' = K''\alpha_{m-1}\alpha_m^{-1}$ and $\alpha_m = \int_0^1\cdots\int_0^1 [0 < (1,x) \leqq 1]dx_0\ldots dx_{m-1}$.

Observing that $\alpha_k = \{k!\}^{-1}$ for $k = m-1$ or m and that the function under the integral in (4) is bounded by $[B_j](x)$, we have that (4) substituted into (3) implies $P_i f^{-1}[B_j] \leqq K'\lambda_m[B_j]$, where $K' = \alpha_{m-1}\,\alpha_m^{-1}\,K'' = m\,K''$, and the lemma is proved.

Lemma 1.2 proves that condition (II") implies condition (II') when $\mu = P_{.}$.

APPENDIX 2

Truncation of $\bar{k}$ to a Convex Set of R^s.

Let $T = \{\tau = (\tau_1,\ldots,\tau_s)|\tau_i \in R\}$ be a convex set of R^s. With T as the nuisance parameter set of Chapter IV, we shall give a constructive method of truncating $\bar{k}(X)$, given by (4.5), to T.

Lemma 2.1.

If τ_0 is an exterior point of T, then there exists a unique point τ_0' in the boundary of T, denoted B(T), such that $\|\tau_0-\tau_0'\| = \min_{\tau\in\bar{T}}\|\tau_0-\tau\|$, where $\bar{T}$ is the closure of the convex set T.

Proof. Since $\bar{T}$ is closed, there exists a $\tau_0' \in \bar{T}$ such that $\|\tau_0-\tau_0'\| = \min_{\tau\in\bar{T}}\|\tau_0-\tau\|$. Suppose τ_0' is an inner point of T. Then the line segment $\lambda\tau_0' + (1-\lambda)\tau_0$, $0 \leq \lambda \leq 1$, would intersect the boundary of T at a point $\tau_0'' = \lambda_0\tau_0' + (1-\lambda_0)\tau_0$, $0 < \lambda_0 < 1$. Then $\tau_0'' \in \bar{T}$ and $\|\tau_0-\tau_0''\| = \lambda_0\|\tau_0-\tau_0'\| < \|\tau_0-\tau_0'\|$, which is a contradiction. Therefore, τ_0' is not an inner point, and hence is a boundary point of T.

To show that τ_0' is unique, suppose there exists τ_1 in the boundary of T such that $\|\tau_0-\tau_1\| = \min_{\tau\in\bar{T}}\|\tau_0-\tau\|$, $\tau_1 \neq \tau_0'$. Then the three points τ_0, τ_0', and τ_1 are the vertices of an isosceles triangle having equal sides $\|\tau_0-\tau_0'\| = \|\tau_0-\tau_1\|$. Hence, the mid-point of the base, given by $\tau_2 = \frac{1}{2}(\tau_0' + \tau_1)$ satisfies the Pythagorean equality

$$(1) \quad \|\tau_1-\tau_2\|^2 + \|\tau_0-\tau_2\|^2 = \|\tau_0-\tau_1\|^2 .$$

But, by convexity of T, we have $\tau_2 \in \bar{T}$ and thus $\min_{\tau\in\bar{T}}\|\tau_0-\tau\| \leq \|\tau_0-\tau_2\| \leq \frac{1}{2}\|\tau_0-\tau_0'\| + \frac{1}{2}\|\tau_0-\tau_1\| = \min_{\tau\in\bar{T}}\|\tau_0-\tau\|$. Hence, (1) implies $\|\tau_1-\tau_2\| = 0$, or $\tau_0' = \tau_1$, a contradiction. Therefore, τ_0' is unique and the lemma is proved.

Lemma 2.2 (Blackwell and Girshick).

Let T be a convex set in R^s. If τ_1 is an inner point of T and τ_2 a boundary point of T, then the points $(1-\lambda)\tau_1 + \lambda\tau_2$ are inner points of T for $0 \leqq \lambda < 1$.

Proof. See Lemma 2.2.1(a) of [1].

With the aid of Lemmas 2.1 and 2.2 we can now truncate $\bar{k}(X) = (\bar{k}_1(X),\ldots,\bar{k}_s(X))$ to T as follows. Let $\tau_0 \varepsilon T$ be a fixed interior point of T, which exists by the assumption on T in Chapter IV. Denote $\bar{k}^*(X) = (\bar{k}_1^*(X),\ldots,k_s^*(X))$ as the truncation of $\bar{k}(X)$ to T given by,

$$(2)\qquad \bar{k}^*(X) = \begin{cases} \bar{k}(X) & \text{if } \bar{k}(X)\ \varepsilon\ T \\ \bar{k}'(X) & \text{if } \bar{k}'(X)\ \varepsilon\ T,\ \bar{k}(X) \not\varepsilon\ T \\ (\lambda_0 N)^{-1}\tau_0+(1-(\lambda_0 N)^{-1})\bar{k}'(X) & \text{if } \bar{k}'(X) \not\varepsilon\ T, \\ & \quad \bar{k}(X) \not\varepsilon\ T, \end{cases}$$

where $\bar{k}'(X)$ is the unique boundary value of T closest to $\bar{k}(X)$ given in Lemma 2.1 and $\lambda_0 = \max_{\tau\varepsilon B(T)}\|\tau_0-\tau\|$. Note that Lemma 2.2 guarantees that $\bar{k}^*(X)\ \varepsilon\ T$ in the case where $\bar{k}'(X) \not\varepsilon\ T$ and $\bar{k}(X) \not\varepsilon\ T$. The truncated estimate $\bar{k}^*$ depends on the fixed value τ_0. Note that from (2) we have that if $\bar{k} \not\varepsilon\ T$, then $\|\bar{k}^*-\bar{k}'\| \leqq (\lambda_0 N)^{-1}\|\tau_0-\bar{k}'\| \leqq N^{-1}$. Thus with T a convex set of R^s we have exhibited a constructive method of truncation meeting the requirements of Chapter IV.

APPENDIX 3

Extension of Results for a Randomized Procedure.

We extend Theorems 2 - 11 to the non-simple, randomized procedure defined by substituting the estimate $\bar{h}$ for $p(\theta)$ (and $\bar{k}$ for τ in Chapter IV) in the simple randomized procedure which assigns equal probabilities of selection among all columns minimizing $(p(\theta), L^{\nu}f)$ in (1.7). Such a randomized, non-simple rule is given by the $N \times n$ matrix of function $T^*(x) = (t^*_{\alpha j}(x))$, where for $j = 0,\ldots,n-1$, $\alpha = 1,\ldots,N$,

(1) $\quad t^*_{\alpha j}(x) = r^{-1}(\alpha,x)$ or 0 according as $j \in$ or $\notin R_{\alpha}(x)$,

where $R_{\alpha}(x) = \{j \,|\, (\bar{h},L^{j}f(x_{\alpha})) = \min_{k}(\bar{h},L^{k}f(x_{\alpha}))\}$, having cardinality $r(\alpha,x)$. We shall show that Theorems 2 - 11 (also, substitute $\bar{k}^*$ for τ in Chapter IV) hold for the randomized procedure $T^*(x)$.

Let $\mathfrak{N}$ be the class of all permutations on the integers $\{0,\ldots,n-1\}$. The elements of $\mathfrak{N}$, denoted by π, are 1-1 functions of $\{0,\ldots,n-1\}$ onto itself defined by $\pi(0,\ldots,n-1) = \{\pi(0),\ldots,\pi(n-1)\}$, where $\pi(j) \in \{0,\ldots,n-1\}$ and $\pi(j) = \pi(k)$ if and only if $j = k$. Let ' denote the identity permutation having $'(j) = j$ for $j = 0,\ldots,m-1$. Now define the following class of non-randomized rules $t^{\pi}_{\bar{h}}$, $\pi \in \mathfrak{N}$, given by

$$(2) \quad t^{\pi}_{\bar{h},j}(x_{\alpha}) = \begin{cases} 1 \text{ if } (\bar{h},L^{j\nu}f(x_{\alpha})) < \text{ or } \leqq 0 \text{ according as} \\ \quad \pi(\nu) < \pi(j) \text{ or } \pi(\nu) > \pi(j) \\ 0 \text{ otherwise.} \end{cases}$$

Note that $t'_{\bar{h},j}(x_{\alpha})$ is that particular non-randomized, non-simple rule given by (1.12) for which Theorems 2-11 are proved. Modifications of this rule were made in Chapters II and IV and the corresponding modifications hold for the permuted rules in (2).

Now average the regret risk functions of the n! rules $t^{\pi}_{\underline{h}}$ and interchange the order of summation and integration to obtain Theorems 2 - 11 holding for the non-simple procedure defined by the N x n functions

$$(3) \qquad (n!)^{-1} \sum_{\pi\varepsilon\mathcal{N}} t^{\pi}_{\underline{h},j}(x_\alpha),\ \alpha = 1,\ldots,N,\ j = 0,\ldots,n-1.$$

We shall now prove that (3) = $t^*_{\alpha j}(x)$.

Fix α, j, x and let $r = r(\alpha,x)$, $R = R_\alpha(x)$. Observe that $j \notin R$ implies $t^{\pi}_{\underline{h},j}(x_\alpha) = 0$ for all $\pi \varepsilon \mathcal{N}$. Hence, if $j \notin R$, (3) = 0 and so is $t^*_{\alpha j}(x)$ given by (1). Next, observe that if $j \varepsilon R$, then $\sum_{\pi\varepsilon\mathcal{N}} t^{\pi}_{\underline{h},j}(x_\alpha) = \sum_{\pi\varepsilon\mathcal{N}} [\pi(\nu) > \pi(j)$ for all $\nu \varepsilon R$, where $\nu \neq j] = \sum_{t=o}^{n-r} \sum_{\{\pi|\pi(j)=t\}} [\pi(\nu) > t$ for all $\nu \varepsilon R$, where $\nu \neq j]$. The number of permutations $\pi \varepsilon \mathcal{N}$ having the permuted position $\pi(j)$ fixed at t and with r-1 permuted positions $\pi(\nu)$ greater than t is $(n-t-1)!\ P(n-r,t)$, where $P(n,k)$ is the permutation of n objects k at a time. With $C(n,k)$ denoting the combination of n objects k at a time, we have $(n-t-1)!\ P(n-r,t)=C(n-t-1,r-1)\ (n-r)!\ (r-1)!$ Hence, by our earlier observations we have that if $j \varepsilon R$, then $\sum_{\pi\varepsilon\mathcal{N}} t^{\pi}_{\underline{h},j}(x_\alpha) = \sum_{t=o}^{n-r}(n-t-1)!\ P(n-r,t) = (n-r)!\ (r-1)! \sum_{t=o}^{n-r} C(n-t-1,r-1)$. Finally, since $\sum_{t=o}^{n-r} C(n-t-1,r-1) = C(n,r)$, (see Feller [3], (12.8), p. 62), we conclude that if $j \varepsilon R$,

$$(4) \qquad (n!)^{-1} \sum_{\pi\varepsilon\mathcal{N}} t^{\pi}_{\underline{h},j}(x_\alpha) = r^{-1} \quad .$$

Hence, we have shown that $t^*_{\alpha,j}(x)$ defined by (1) equals (3). Since Theorems 2 - 11 hold for the procedure given by (3), the same is true for $T^*(x)$ defined by (1).

BIBLIOGRAPHY

[1] BLACKWELL, DAVID and GIRSHICK, M. A. (1954). Theory of Games and Statistical Decisions. Wiley, New York.

[2] CRAIG, CECIL C. (1936). On the Tchebycheff inequality of Bernstein. Ann. Math. Statist. **4**, 94-102.

[3] FELLER, WILLIAM (1957). An Introduction to Probability Theory and Its Applications. **1** (2nd ed.) Wiley, New York.

[4] HALMOS, PAUL R. (1950). Measure Theory. Van Nostrand, Princeton.

[5] HANNAN, JAMES F. (1953). Asymptotic solutions of compound decision problems. Institute of Statistics Mimeograph Series, Univ. of North Carolina. No. 68, 1-77.

[6] HANNAN, JAMES F. (1957). Approximation to Bayes risk in repeated play. Contributions to the Theory of Games, **3**, 97-139. Ann. of Math. Studies No. 39, Princeton Univ. Press.

[7] HANNAN, JAMES F. and ROBBINS, HERBERT (1955). Asymptotic solutions of the compound decision problem for two completely specified distributions. Ann. Math. Statist. **26**, 37-51.

[8] HANNAN, JAMES F. (1957). Unpublished lecture notes. Michigan State Univ.

[9] LOÈVE, MICHEL (1960). Probability Theory. (2nd ed.) Van Nostrand, Princeton.

[10] ROBBINS, HERBERT (1951). Asymptotically subminimax solutions of compound statistical decision problems. Proc. Second Berkeley Symp. Math. Statist. Prob. 131-148. Univ. of California Press.

[11] SAMUEL, ESTER (1961). On the compound decision problem in the nonsequential and the sequential case. Ph.D. Thesis, Columbia Univ.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to Professor J. F. Hannan for suggesting the area of research and for his inspiring guidance throughout this investigation. His comments aided greatly in improving and simplifying many earlier results. I am deeply indebted to him for giving me the unpublished lecture notes upon which section 1.3 is based.

Many thanks are also due to Professor G. Kallianpur, who, in the absence of Professor Hannan, served as my guide. His suggestion of the problem of Chapter IV, as well as many illuminating discussions, were extremely helpful.

The financial support and creative atmosphere provided by the Department of Statistics at Michigan State University, and by the Applied Mathematics Division of Argonne National Laboratory, were invaluable. This research was supported in part by the National Science Foundation and the Office of Naval Research. I am also grateful for support of graduate study through National Science Foundation fellowships from June, 1960 to September, 1961.

Finally, I wish to thank Dr. R. F. King for his editorial help, and M. Jedlicka and G. Krause for their excellent typing and cheerful attitude in preparation of the manuscript.